

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES A. MATHEMATICAL AND PHYSICAL SCIENCES

VOL 203


LONDON

Published for the Royal Society by the
Cambridge University Press
Bentley House, N.W.1

24 October 1950

Printed in Great Britain at the University Press, Cambridge
(Brooke Crutchley, University Printer)
and published by the Cambridge University Press
Cambridge, and Bentley House, London
Agents for Canada, India, and Pakistan: Macmillan

CONTENTS

SERIES A VOLUME 203

No. A 1072. 7 September 1950

	PAGE
Bakerian Lecture. Physics above 20,000 kg./cm. ² By P. W. Bridgman, For. Mem. R S. (Plate 1)	1
Kinetic salt effects on the inversion of sucrose. By E. A. Guggenheim, F.R.S. and L. A. Wiseman	17
The propagation of a sound pulse in the presence of a semi-infinite, open-ended channel. II. By W. Chester	33
Measurement of the adsorption of surface-active agents at a solution/air interface by a radiotracer method. By D. J. Salley, A. J. Weith, Jr., Ann A. Argyle and J K. Dixon	42
Diffusion and flow of gases and vapours through micropores. I. Slip flow and molecular streaming. By P. C. Carman	55
The theory of the transport phenomena in metals. By E. H. Sondheimer	75
The surface impedance of superconductors and normal metals at high frequencies. IV. Impedance at 9400 Mc./sec. of single crystals of normal and superconducting tin. By A. B. Pippard	98
Liquid helium films. I. The thickness of the film. By K. R. Atkins	119
The system silver-magnesium-antimony, with reference to the theory of alloy formation. By B. R. T. Frost and G. V. Raynor	132

No. A 1073. 22 September 1950

The fine structure of atmospheric turbulence in relation to the propagation of sound over the ground. By E. G. Richardson. (Plates 2 and 3)	149
Diffusion and flow of gases and vapours through micropores. II. Surface flow. By P. C. Carman and P. le R. Malherbe	165
On the atomic theory of elasticity. By Kun Huang	178
The surface impedance of superconductors and normal metals at high frequencies. V. Analysis of experimental results for superconducting tin. By A. B. Pippard	195
Field variation of the superconducting penetration depth. By A. B. Pippard	210
Size effect variation of the electrical conductivity of metals. By D. K. C. MacDonald and K. Sarginson	223
Liquid helium films. II. The flow of the film. By K. R. Atkins	240
Stability of viscous flow over concave cylindrical surfaces. By D. Meksyn	253
On the Bose-Einstein condensation. By S. R. de Groot, G. J. Hooyman and C. A. ten Seldam	266

No. A 1074. 10 October 1950

PAGE

The Bell Telephone Laboratories—an example of an institute of creative technology. By Mervin J. Kelly. (Plates 4 to 6)	287
Studies in copolymerization. The evaluation of the kinetic coefficients for the system styrene-butyl acrylate. By E. J. Arlman and H. W. Melville, F.R.S.	301
Polypeptide chain configurations in crystalline proteins. By Sir Lawrence Bragg, F.R.S., J. C. Kendrew and M. F. Perutz	321
The decay of axisymmetric turbulence By S Chandrasekhar, F.R.S.	358
Polarizability and internuclear distance in the hydrogen molecule and molecu- lion. By R. P. Bell, F.R.S. and D. A. Long	364
The nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K. I. Thermodynamic measurements made from a study of the field-dependence of the adiabatic susceptibility. By C. G. B. Garrett	375
The nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K. II. The nuclear specific heat. By C. G. B. Garrett.	392
Emission of heavy fragments in nuclear explosions. By D. H. Perkins. (Plates 7 to 10)	399
The derivation of a model solar chromosphere from radio data. By J. H. Piddington	417

No. A 1075. 24 October 1950

Entropies of mixing in certain athermal binary solutions. By E. A. Guggenheim, F.R.S. and M. L. McGlashan.	435
A study of sensitized explosions. VIII. Experimental work on the hydrogen- oxygen reaction sensitized by chloropicrin. By P. G. Ashmore and R. G. W. Norrish, F.R.S.	454
A study of sensitized explosions. IX. Experimental work on the sensitization of ignitions of hydrogen-chlorine mixtures by chloropicrin. By P. G. Ashmore and R. G. W. Norrish, F.R.S.	472
The kinetics of the thermal decomposition of normal paraffin hydrocarbons. III. Activation energies and possible mechanisms of molecular reactions. By K. U. Ingold, F. J. Stubbs and Sir Cyril Hinshelwood, F.R.S.	486
Eigenfunction problems with periodic potentials. By E. C. Titchmarsh, F.R.S.	501
The cataphoresis of spherical, solid non-conducting particles in a symmetrical electrolyte. By F. Booth	514
The electroviscous effect for suspensions of solid spherical particles. By F. Booth	533
Exact solutions for flow of a perfect gas in a two-dimensional Laval nozzle. By T. M. Cherry	551
The propagation of spherical blast. By G. B. Whitham	571
The application of phase-contrast to the ultra-violet microscope. By E. W. Taylor. (Abstract)	582
Index	583

24 OCT 1950

BAKERIAN LECTURE

Physics above 20,000 kg./cm.²

BY P. W. BRIDGMAN, FOR. MEM. R S.

Harvard University, U.S.A.

(Delivered 4 May 1950—Received 4 May 1950)

[Plate 1]

Pressures up to the general order of 20,000 kg./cm.² may be obtained and their effects measured in vessels made of single blocks of the best grades of steel. There is no sharp limit to the pressures which may be so obtained, for the life of the apparatus decreases rapidly in the 20,000 region, so that the upper limit is dictated by considerations of economy in conjunction with the significance of the phenomenon under investigation. The absolute limit for such mono-bloc apparatus for even a single application of pressure is, in my own experience, something less than 30,000 kg./cm.². To reach higher pressures some different principle of construction must be employed. In this lecture I propose to review some of the technical problems which arise in constructing such apparatus and to indicate some of the experimental results in the pressure range above 20,000 kg./cm.².

In the construction of apparatus for this range it is almost necessary to afford external support for the vessel in such a way that the support becomes more effective the higher the internal pressure. This rules out for our consideration apparatus of the conventional built-up construction, in which the support is supplied by shrunk-on hoops, or in similar ways, as by wire winding as in a gun. Furthermore, there is no essential advantage in such built-up construction, for a true mono-bloc apparatus can be thrown into a state of stress similar to the built-up apparatus by a preliminary stretching beyond the elastic limit. All my mono-bloc apparatus does receive such a preliminary treatment. In fact, the distribution of stress so produced in a mono-bloc apparatus is somewhat more effective than that in the equivalent composite apparatus, because the distribution of stress is uniform. The built-up or the 'auto-frettagged' mono-bloc apparatus reaches an obvious upper limit when the initial compressive stress at the inner surface exceeds the critical plastic flow stress.

A simple method of giving support to the external curved surface of a pressure vessel is shown in figure 1. The vessel is given the shape of a cone fitting into a similar conical cavity in a massive block. Pressure is produced in the interior in the conventional way by the thrust of a piston. It will be seen that the geometry is such that the same thrust of the piston which produces the internal pressure also drives the vessel as a whole into its conical seat, and in so doing produces an external supporting pressure which increases proportionally to the thrust. By making the external angle of the cone more acute the ratio of external pressure to

P. W. Bridgman

thrust may be increased indefinitely. Simple considerations show that, neglecting friction, the intensity of external pressure is equal to the total thrust divided by the area of the annulus the inner and outer radii of which are the radii of the two ends of the conical frustum of the vessel where it enters and leaves the block.

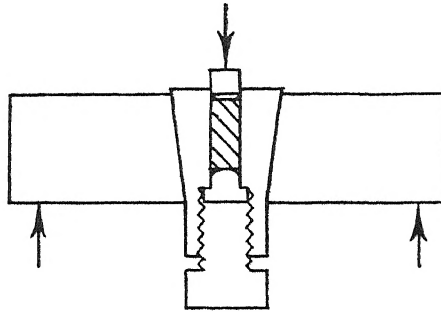


FIGURE 1. Illustrating the conical support by which the external pressure on the pressure vessel increases proportionally to the internal pressure.

There are two limits to the intensity of the external supporting pressure that can be reached with this design. One limit is set by the internal pressure which the supporting block will tolerate. With blocks of practical dimensions this limit is between 15,000 and 20,000 kg./cm.². It may be increased by making the supporting block itself multiple, with one conical sleeve sliding inside another, with the angles so graded that the successive internal pressures progress smoothly. When using more than one stage of support the limitations on the geometry imposed by friction at the sliding surfaces have made it necessary to provide the longitudinal thrust on the vessel independently of the piston. In practice I have not found it necessary to use more than two stages of such a multiple construction, because the limit set by the other effect is presently reached. This is an extrusion of the whole pressure vessel through the conical support by the longitudinal thrust.

The upper limit to which I have used vessels constructed with external conical support is between 50,000 and 60,000 kg./cm.². It was obvious that the upper limit possible with the particular geometrical design was reached because not infrequently fracture occurred on systems of three mutually perpendicular planes. It might be possible to improve the geometry somewhat, but no serious investigation was made of this possibility because another limit is reached at the same time as the limit on the strength of the containing vessel. This is the limit imposed by the strength of the piston with which pressure is produced. This limit had already been passed for steel pistons, the upper limit of which is in the neighbourhood of 30,000 kg./cm.². Most fortunately carboly, which is tungsten carbide cemented with cobalt and which is now much used for tools, is approximately twice as strong in compression as the best steel, and has proved to be a suitable material for pistons up to somewhat beyond 50,000 kg./cm.². The upper limit which I have found for the strength of a carboly piston is somewhat less than 70,000 kg./cm.², but it is desirable not to try for the limit of strength in the piston because carboly is capricious in its upper range.

To reach pressures beyond the 50,000 limit a more effective method of support must be found for the vessel, and at the same time some method must be found for strengthening the piston. Both ends may be attained simultaneously by immersing the entire pressure apparatus, vessel and piston together, in a fluid carrying a high hydrostatic pressure. Theoretically this type of construction could be multiplied indefinitely, constructing one pressure apparatus inside another, the inner apparatus beginning where the outer leaves off. The practical difficulties in such an arrangement are obvious in finding a suitable method of manipulating the inner pieces of apparatus. These practical complications have up to the present limited the application of this multiple or 'cascade' principle to only a single stage. It has now been found possible to manipulate a diminutive pressure apparatus inside an outer apparatus in which there is a pressure of from 25,000 to 30,000 kg./cm.². It is possible in this way to reach in the inner apparatus pressures of 100,000 kg./cm.² or more and to study some of the effects of such pressures. The increase in range, from 30,000 to 100,000 is more than would be suggested by a literal application of the cascading principle, which would suggest an extension from 30,000 to 60,000. The greater than theoretical increase is possible because of the change in the intrinsic properties of many materials brought about by a pressure of the order of 25,000 kg./cm.². Materials such as steel become indefinitely plastic under such pressures, and at the same time support a greatly increased work hardening so that they will support greater stresses than normal. Carboloy loses its brittleness, so that it will support greatly increased compressive stresses, to which it yields slowly and plastically without abrupt fracture. Carboloy also increases in tensile strength by a factor of as much as three, at the same time losing so much of its brittleness in tension that it is no longer troublesome. This makes it possible to make the diminutive pressure vessel for reaching 100,000 of carboloy. Along with the increased strength there is another advantage in constructing the high-pressure vessel of carboloy in that the elastic distortion of carboloy is only about one-third that of steel. The elastic distortion of a steel vessel under such pressures is so high as to make it very difficult to manage, the increase of internal area of a steel vessel exposed to an internal pressure 100,000 kg./cm.² being of the order of 10 %.

Utilizing these two types of support, conical support and complete support by a liquid, various sorts of apparatus have been constructed for use in different ranges, and different sorts of physical phenomena have been investigated with them. The first range is to 30,000 kg./cm.². An apparatus for this pressure range is shown in figure 2 and figure 3, plate 1. The pressure vessel has an inner diameter of 0.5 in. and an effective volume of the order of 15 cm.³. Pressure is transmitted with a true liquid, pentane, which just escapes freeing under a pressure of 30,000 kg./cm.² at room temperature. Electrically insulated leads may be taken into the chamber, and means have been found for getting seven such leads into the vessel through a single plug. This allows considerable flexibility in the sort of measuring devices that may be mounted within the chamber, and it is possible to set up within the pressure vessel the equivalent of a miniature laboratory and conduct a wide variety of measurements. In fact, there is no reason at all why all the measurements which

I formerly made to 12,000 kg./cm.² should not be extended to 30,000, and this programme is now partly under way.

Measurements in the 30,000 range are possible of a rather good degree of precision, of the order of 0.1 %. In order to attain this precision there must in the

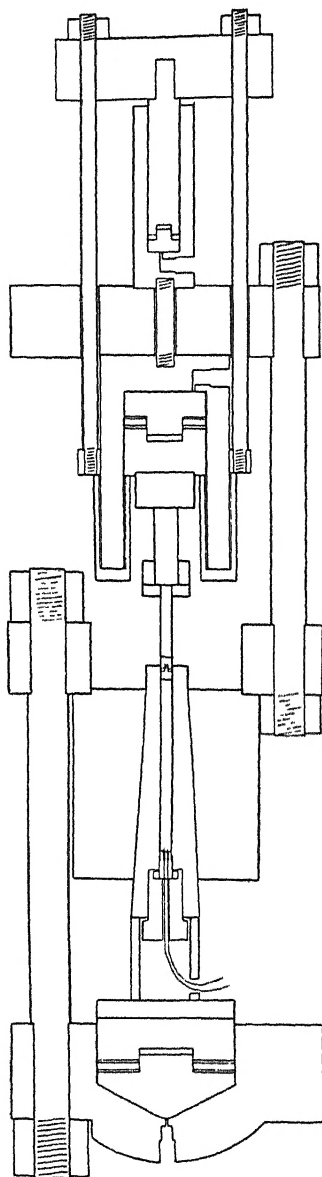


FIGURE 2. Section of the general assembly of the apparatus for producing 30,000 kg./cm.².

first place be pressure fixed points, established with sufficient accuracy. In the range to 12,000 a single fixed point, the freezing of mercury at 0° C at 7,640 kg./cm.², was adequate. The ultimate reference here was to a free piston gauge, of special design. With this free piston gauge it was established that the increase of electrical resistance of

manganin under pressure is linear up to 12,000 kg./cm.² within 0.1 %. The linearity having been established, the manganin gauge could be used as a secondary pressure-measuring device, and its constant could be determined by a single calibration against mercury freezing under pressure at 0° C. In extending the range from 12,000 to 30,000 the first question that had to be answered was how nearly the manganin gauge is linear over the wider range. This required the finding of new standard pressures of reference in the new range. The newly discovered transition of bismuth near 25,000 presented itself as eminently suitable, since the transition runs cleanly, with little overshooting, and the volume change is large enough to make the manipulations easy. To establish the exact pressure of this transition required a special procedure. The free piston gauge I do not believe can be used to pressures materially higher than my previous limit at 13,000 kg./cm.², because of rapidly increasing difficulties due to viscosity of the transmitting medium and the magnitude of the forces. At the same time, some method of referring the calibration to the definition of pressure in terms of force divided by area seemed unavoidable. The calibration was affected by dispensing with the use of any liquid at all, transmitting the compressive stress directly to the bismuth with a steel piston, making the ratio of thickness of the bismuth to its diameter as small as possible in order to avoid error from friction, and determining the distortion of the cross-section under the internal pressure by auxiliary experiments in which the vessel was filled with liquid and the change of cross-section directly measured by a suitable device mounted within the vessel. In this way the bismuth fixed point in the neighbourhood of 25,000 was established with an accuracy of the order of 0.1 %. This point having been fixed, the deviation from linearity of the manganin gauge over the greater range was then determined. The correction fortunately turned out to be small, the error in extrapolating linearly from the mercury fixed point to 30,000 being only about 1 %.

The manganin gauge remains the most convenient secondary pressure gauge up to 30,000 kg./cm.². The most accurate work demands a calibration of each individual gauge at two fixed points, those afforded by mercury at 0° C and by bismuth at room temperature. For much high-pressure work it is accurate enough to calibrate at only a single point, which may be conveniently taken as the freezing pressure of mercury at room temperature instead of at 0° C.

It is feasible to construct much larger apparatus for the 30,000 range than that just described. One such apparatus has been constructed in an investigation of the possibility of the diamond synthesis. The piston diameter was 2 in. and the effective length of the high-pressure chamber was 6 in. The piston was actuated by an industrial 1000-ton press. There is no apparent reason why still larger apparatus should not be constructed.

The second pressure range is up to 50,000 kg./cm.². Such an apparatus is shown in figure 4 and figure 5, plate 1. It is no longer feasible to transmit pressure with a liquid, since all practical liquids have frozen. Even the permanent gases argon and nitrogen, extrapolating by the formulas of Simon, would be solid at room temperature at 50,000 kg./cm.². The only substances which remain fluid at room temperature at 50,000 kg./cm.² are hydrogen and helium. Hydrogen is absolutely

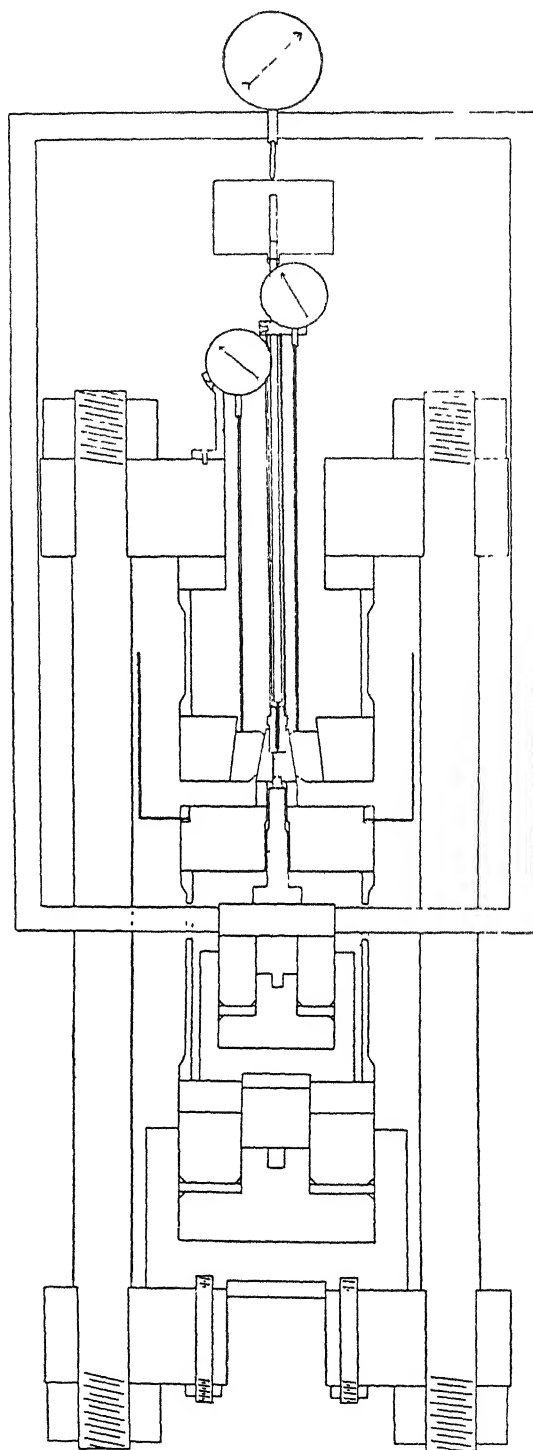


FIGURE 4. Section of the general assembly of the apparatus for producing 50,000 kg./cm.².

ruled out because it attacks steel under pressure, and the technical difficulties in handling helium make it prohibitive. Pressure must therefore be transmitted by a solid—some soft metal is most suitable for this purpose. It might perhaps be thought that some soft organic solid such as rubber or paraffin would be suitable, but such solids stiffen enormously under pressure and become more rigid than steel itself. But the effect of pressure in stiffening a metal, especially one which has a simple atom, is much smaller, and it is possible to transmit pressure with lead, or better, the softer indium, without too great departure from a hydrostatic condition in chambers with a ratio of length to diameter of the order of 2 or 3. Special measurements showed that the shearing strength of these metals under a mean hydrostatic pressure of 50,000 kg./cm.² is only a few hundred kg./cm.², so that the departure from hydrostatic conditions is only a fraction of 1 %.

It will be realized from figure 5 that the size of the pressure vessel for 50,000 is much less than for 30,000. The internal diameter is only 0.25 in. and the capacity is of the order of 0.6 cm.³. The pressure in this apparatus is determined from the total thrust on the piston and the cross-sectional area. The effect of friction is minimized by using the mean of readings taken with increasing and decreasing pressure. The correction for the distortion of the cross-section by internal pressure cannot be determined by independent experiment, but is calculated by elasticity theory. The correction may be checked by comparing results with those obtained in the 30,000 range. The pressures obtained in this way are probably accurate to within about 2 %. This apparatus has up to the present been mostly used in making measurements involving changes of volume. Such measurements include direct measurements of compressibility, and also phase changes, both meltings and polymorphic transitions. It is an advantage of this apparatus, as compared with that for the 30,000 range, that measurements over a temperature range are easily combined with the pressure measurements, whereas with the 30,000 apparatus measurements at other than room temperature are inconvenient, although such have been made up to 100° C. Electrical measurements are more difficult with the 50,000 apparatus. It is possible, however, to get a single insulated lead into the chamber with fairly high current-carrying capacity. With this a few measurements have been made of the resistance of semi-conductors up to 50,000 kg./cm.².

Brief mention may be made of another apparatus good to 40,000 kg./cm.². The working cylinder of this apparatus is carbony, in a shrunk-on steel jacket. Supporting pressure is applied in the conventional way by cones. By giving the cones suitable angles it is so arranged that the change of cross-section vanishes, and the only correction to the apparent change of volume as measured by the displacement of the pistons arises from the shortening of the pistons, which is readily corrected for and is only 2 %. With this apparatus measurements of considerable accuracy may be rapidly made of the changes of volume of solids and of liquids also, the latter by sealing inside a lead capsule, over the complete pressure range to 40,000 kg./cm.². Several hundred substances have been so measured.

The diminutive apparatus for reaching 100,000 kg./cm.² is shown in figure 6. As already stated, this is mounted inside the vessel of the 30,000 apparatus and is completely supported externally by a liquid bearing a pressure from 25,000 to

30,000 kg./cm.². Figure 7 shows the 100,000 apparatus within the 30,000 vessel. At low pressures the 100,000 apparatus is out of action. By properly choosing the amount of liquid in the 30,000 vessel it may be arranged so that the $\frac{1}{2}$ in. piston of the 30,000 vessel comes into contact with the 100,000 pistons when a pressure of 25,000 has been reached. From here on pressure rises in both vessels. Since the

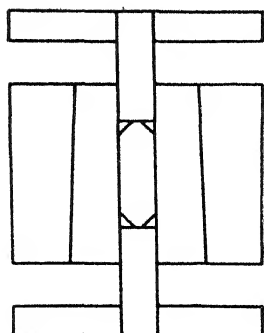


FIGURE 6. The miniature pressure vessel for producing 100,000 kg./cm.². This is mounted inside the 30,000 apparatus. The diameter of the carboloy pistons is $\frac{1}{16}$ in.

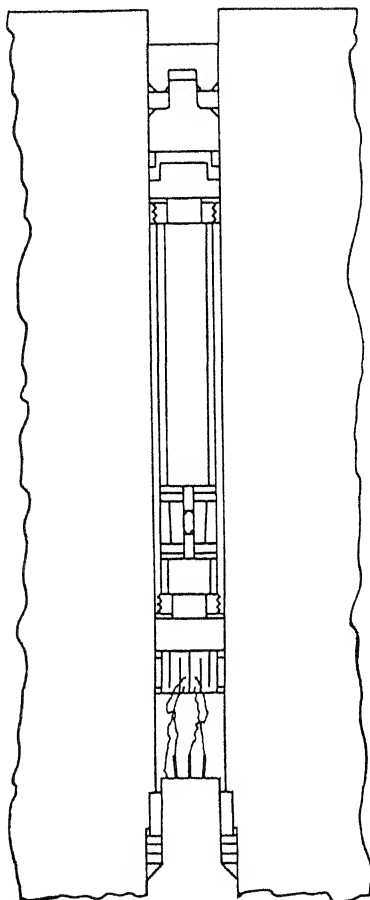


FIGURE 7. Details of the mounting of the 100,000 vessel of figure 6 within the 30,000 vessel.

100,000 vessel is much smaller than the 30,000 vessel, a given piston motion produces a much greater rise of pressure in it, and with the dimensions indicated in the figure pressure rises to 100,000 in the inner vessel before it has reached 30,000 in the outer. With this apparatus changes of volume may be measured as a function of pressure. The pressure is given by the thrust on the piston and the area. The area changes with pressure and a correction has to be applied for the change. Fortunately, the elastic constants of carboloy are three times higher than those of steel, so that the

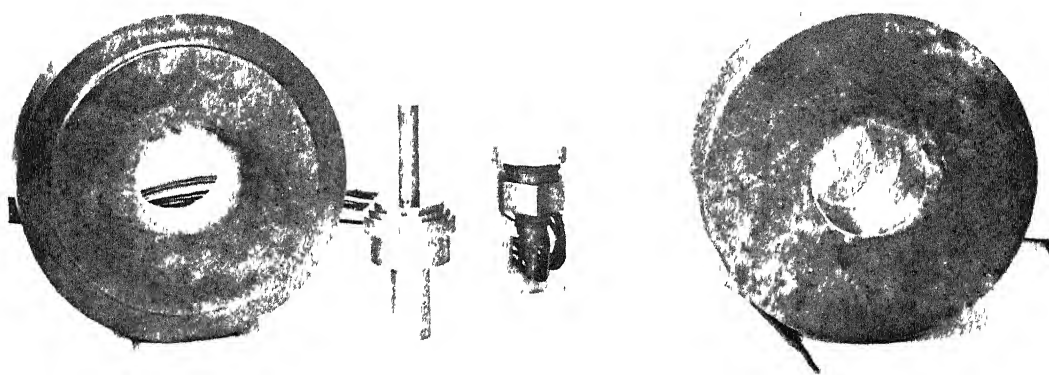


FIGURE 3. The conical pressure vessel for 30,000 kg/cm², the carboloy piston, and the three external reinforcing rings.

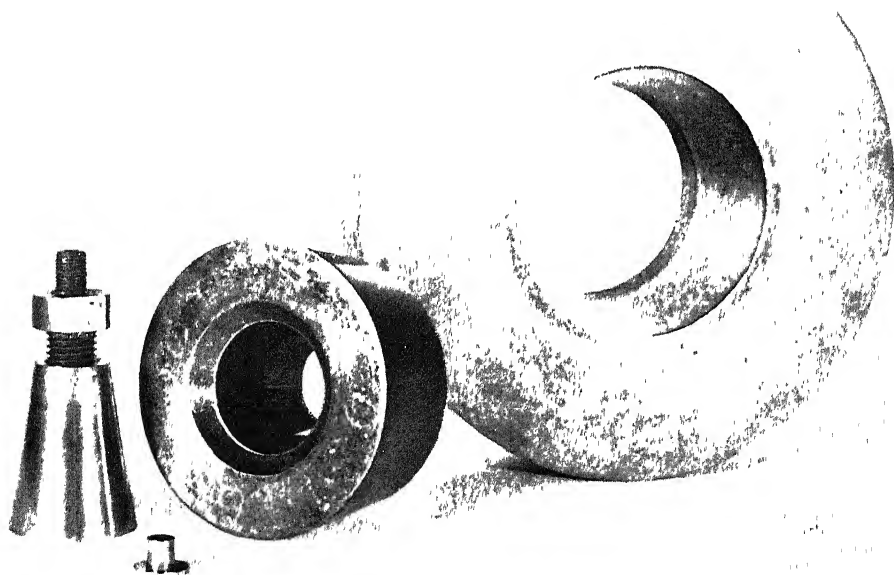


FIGURE 5. The conical pressure vessel for 50,000 kg/cm², the carboloy piston ($\frac{1}{4}$ in. in diameter), and the double external reinforcing rings.

(Facing p. 8)

change of area is small, 3 % at the maximum, and I have contented myself with calculating the correction by the theory of elasticity. The thrust on the piston could in principle be measured by determining the extra thrust on the 30,000 piston over and above that required to produce the known pressure in the 30,000 vessel in the absence of the inner vessel. Such a determination, however, would suffer from so much error from the large friction as to be quite impractical. It is necessary to measure the thrust on the small pistons from inside the pressure chamber. This was accomplished by supporting the 100,000 vessel on a 'grid' which receives the full thrust. This 'grid' is a thin shell of hardened steel, strong enough to withstand the thrust, and so cut by a succession of staggered slots as to have a comparatively high electrical resistance. The electrical resistance changes with the thrust, and by measuring the resistance the thrust may be determined. The thrust coefficient of the grid is itself a function of the pressure in the surrounding liquid. Two ways were found of determining this coefficient which need not be described here. Since the two methods were in full agreement no error may be feared. The motion of the pistons which determine the changes of volume may be measured with sufficient accuracy for many purposes merely by measuring the external motion of the piston of the 30,000 apparatus. However, there are various minor errors in the motion so determined, arising from friction and hysteresis in the external vessel as it slides back and forth in its supporting cones, and for the greatest accuracy the motion of the pistons of the 100,000 vessel must also be measured from within the pressure vessel. It is not too difficult to do this by attaching the pistons to a wire sliding over a contact and measuring the changes of resistance with a potentiometer. Altogether the seven leads of the plug are all occupied: three (with earth) for the grid, three (with earth) for the piston displacement, and one for the manganin gauge.

We now turn from these technicalities connected with the construction of the apparatus and the measurement of pressure to consider some of the phenomena in the range above 20,000. Most investigated have been changes of volume, including in the first place the simple volume compression of single phases. In figure 8 are shown the volume changes of some of the more compressible elements up to 100,000 kg./cm.². Except for the solidified permanent gases, which have not been measured over any extensive pressure range, the most compressible elements are the alkali metals and the alkali earths. The compressibility increases with increasing atomic weight in families of similar elements, which means that caesium is the most compressible metal, as shown in the figure. At 100,000 kg./cm.². the volume of metallic caesium is reduced to approximately three-eighths of its original value. The reason for the increasing compressibility with increasing atomic weight in similar families is that the major part of the volume change of solids in this pressure range is afforded by the compression of the atoms themselves, and the heavier atoms with their more complex electronic structure are more deformable. The range of compressibility of solids through the periodic table is large; the least compressible element is carbon in the form of diamond. Linear extrapolation of measurements of Adams & Williamson (1923) in a low-pressure range would indicate that at 100,000 kg./cm.² the volume change of diamond is less than 1.8 %. All

the elements shown in figure 8 are crystalline except selenium, the volumes of this are for the glassy amorphous modification. Compared with the crystalline elements this shows a characteristic difference, the initial compressibility being relatively high, with a relatively rapid drop of compressibility at higher pressures. This is shown in the figure by the crossing of the curves for barium and selenium and lithium and selenium. The reason for the relatively high initial compressibility of the glassy selenium and its relatively rapid drop at higher pressures is that initially there are comparatively large empty spaces between the molecules because of their haphazard arrangement, and at higher pressures the empty spaces get squeezed out and the atoms forced into more effective contact. The same phenomenon is shown much more strikingly by the true liquids. Initially all liquids have a relatively high compressibility, ten to one hundred times that of the common solid elements. This high initial compressibility is rapidly lost, however, as the vacant spaces between the molecules are squeezed shut. For example, the compression curve of methyl alcohol rises initially twice as rapidly as that of caesium, but in the neighbourhood of 16,000 kg./cm.² it crosses it and near 37,000 crosses that of rubidium. At 50,000 kg./cm.², which is the limit of the measurements for liquids, the compression of methyl alcohol is almost exactly 0.4. This is also very approximately the volume compression under 50,000 of nearly all common liquids, including organic compounds and such other common liquids as water and carbon bisulphide, but excluding liquid metals like mercury. The solid phase of compounds normally liquid is always less compressible than the liquid phase at the same pressure and temperature, but the difference is not large. It is a consequence that the volume decrement between 25,000 and 50,000 kg./cm.² of the solid phase of the common liquids varies over a surprisingly small range, how small is shown in table 1. For comparison in the same table are shown the volume decrements in the same pressure range for those substances which remain liquid in the subcooled condition at these pressures at room temperature. Again the variation in numerical magnitudes is small. In both cases, for solids and liquids, the compression in this pressure range must be essentially that of the molecules themselves, which means the sum of the compressibilities of the constituent atoms. This small compressibility range must be more or less fortuitous in view of the range of compression of all the elements, but it would seem indicated that the atomic compression of the constituents of the common compounds, that is, hydrogen, carbon, oxygen, nitrogen, chlorine, bromine and sulphur, do not differ greatly from each other.

Perhaps the most dramatic example of a high initial compressibility at low pressures and its exhaustion at higher pressures due to removal of the empty spaces between the structural units is afforded by quartz glass. Quartz glass is one of a few abnormal substances the compressibility of which increases with increasing pressure. The curve of volume versus pressure is therefore concave toward the pressure axis. The concavity continues up to 35,000 kg./cm.², the slope of the curve continually increasing. At 35,000 the concavity abruptly ceases, there is a cusp in the curve of volume versus pressure, and at pressures beyond 35,000 the curve is convex and the compressibility decreases as is normal with rising pressure. Qualitatively, the effects are as if the quartz glass initially contained lenticular cavities, which with

increasing pressure are squeezed flatter and flatter and thus offer continually decreasing resistance to further compression, until at 35,000 the opposite sides of the lenticular cavities are squeezed into contact, and the effect abruptly ceases.

In figure 8 the curves for caesium, barium, bismuth and antimony contain step-like discontinuities; these are due to polymorphic transitions. Of these four elements bismuth has the largest number of transitions, actually five instead of the four shown in the figure, the lower large transition being actually two transitions so closely spaced that they could not be resolved in the figure. Bismuth thus has six known modifications. There is a very close parallelism between these modifications of bismuth and those of ice. The number of known stable modifications is the same for both substances, and, furthermore, both bismuth and ice are abnormal in that the solid stable at atmospheric pressure is less dense than its liquid. Antimony is known to be similar to bismuth in its crystal structure. It is therefore natural to suppose that the transition of antimony at 85,000 kg./cm.² corresponds to the first of the two transitions at 25,000.

TABLE 1. VOLUME DECREMENTS, IN FRACTIONAL PARTS OF INITIAL VOLUME, BETWEEN 25,000 AND 50,000 KG./CM.²

solids		liquids	
methyl alcohol	0.053	methyl alcohol	0.054
ethyl alcohol	0.057	<i>n</i> -propyl alcohol	0.067
<i>n</i> -butyl alcohol	0.053	<i>i</i> -propyl alcohol	0.066
ethyl bromide	0.059	<i>n</i> -amyl alcohol	0.063
<i>n</i> -propyl bromide	0.045	<i>n</i> -butyl bromide	0.070
ethyl acetate	0.054	ethyl acetate	0.061
<i>n</i> -amyl oxide	0.059		
chloroform	0.054		
carbon bisulphide	0.059		
benzene	0.055		
chlorobenzene	0.054		
methylene chloride	0.051		
ethylene bromide	0.055		
cyclohexane	0.052		
methyl cyclohexane	0.051		
<i>p</i> -xylene	0.045		
water	0.047		

In addition to the polymorphic transitions shown in figure 8 a large number of others have been studied, most of them over a range of temperature as well as of pressure, so that the complete phase diagrams are known and the thermodynamic parameters of the transitions are completely established. Polymorphism proves to be a very common phenomenon. Approximately one-third of all substances examined in the temperature range up to 200° C exhibit polymorphism, and in many cases the new polymorphic forms exist only in the high-pressure domain. Under the very wide ranges of temperature and pressure encountered in the crust of the earth it would seem probable that most substances exist in crystalline modifications not known in the laboratory. Since polymorphic changes are in

general accompanied by volume changes, the importance of this for geophysics is obvious.

The transition of caesium at 45,000 kg./cm.² shown in figure 8 is in all probability of a different type from the others hitherto considered. All these other transitions, including the transition of caesium at 23,000, doubtless correspond to a change of lattice. There is theoretical support for this in the case of the transition of caesium at 23,000. At atmospheric pressure caesium crystallizes in the body-centred cubic arrangement. Bardeen (1938) has succeeded in calculating the pressure-volume

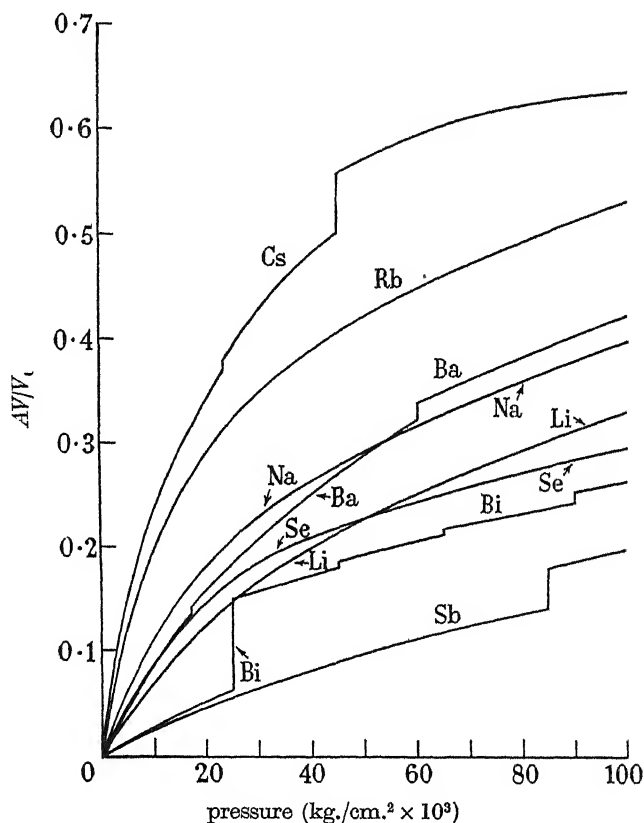


FIGURE 8. The relative volume decrements of a number of elements to 100,000 kg./cm.² at room temperature. The discontinuities are polymorphic transitions.

curves for all the alkali metals with a gratifying agreement with experiment. In the course of the calculation he found that caesium should have a reversible transition to the face-centred cubic arrangement, and predicted the occurrence of the transition at the pressure indicated, 23,000, before the existence of the transition had been certainly established by experiment. Now the face-centred cubic arrangement is close-packed, so that it would seem that no other lattice is to be expected at still higher pressures. The transition at 45,000, with its large change of volume of 16 %, would seem therefore to be a consequence of some electronic rearrangement within the atom. This surmise was later checked by detailed calculation by Sternheimer

(1949), who found that a change of an electron from a $6s$ to a $5d$ orbit is in accord with all the experimental data. Recently, another probable example of the same sort of thing has been found. Several years ago I found a transition in cerium at $12,000 \text{ kg./cm.}^2$ with a volume change of 16 %, practically the same as the change of volume of caesium. Recently at the University of Chicago (Lawson & Ting-Yuan Tang 1949) X-ray photographs have been obtained by a most ingenious technique of the high-pressure modification while under pressure, and it has been established that there is no change of lattice at the transition. In this case also detailed calculation has shown the probable change of electronic orbit to account for the transition.

Caesium and cerium suggest what may be expected in other cases in the domain of pressure of the order of millions or tens of millions of atmospheres. At still higher pressures theory indicates that all atoms break down and their behaviour under pressure becomes simple, approaching that of a perfect gas. In the range up to $100,000 \text{ kg./cm.}^2$ this ultimate behaviour is hardly even foreshadowed. There have been a number of attempts to extrapolate the experimental data into the transitional zone and beyond by making some sort of smooth connexion between the experimental results and the presumptive ultimate simple behaviour of a perfect gas. The results for caesium and cerium would suggest that it may well be that the curves in the transitional zone are not smooth, but may be full of discontinuous incidents as different electronic orbits successively break down or are rearranged.

All the transitions considered up to now, whether due to changes of the lattice or to electronic rearrangements, are reversible. In addition, there are irreversible transitions produced by pressure. Black phosphorus was an early example. Black phosphorus is formed irreversibly from yellow phosphorus by exposing it to pressures above $12,000 \text{ kg./cm.}^2$ at temperatures above 200°C . More recently, it was found that red or violet phosphorus may be irreversibly transformed to black at room temperature by pressures of $50,000$ or higher. Black phosphorus is indefinitely stable under atmospheric conditions with respect to other modifications, and has radically different physical properties, such as electrical conductivity. Another example is carbon bisulphide. At pressures above $40,000 \text{ kg./cm.}^2$ and a temperature of 175°C or higher this is irreversibly changed to a permanent black solid. Neither of these irreversible transitions was predicted, and even yet there seems to be no adequate theoretical method of predicting when such changes are to be expected. Until we have command of some method of prediction the possibility must be kept in mind that there may be other common substances which can be permanently changed into new forms with radically different properties by the application of a pressure sufficiently high to exceed some threshold value.

We next briefly consider another sort of effect produced by pressure, namely the change in the viscosity of liquids. Studies already made of the effect of pressures up to $12,000 \text{ kg./cm.}^2$ had shown that in general the effect is very large, and that instead of tapering off at high pressure it becomes more accentuated. If the logarithm of the viscosity is plotted against pressure it will be found that in general at low pressures the curve is concave toward the pressure axis, but there is a point of reversal, and at the upper end of the range the curve becomes convex toward the pressure axis, indicating an eventual increase of viscosity more rapid than ex-

ponential. The magnitude of the effect is a strong function of the complexity of the molecule, as is also the pressure of reversal of curvature, which occurs at higher pressures the simpler the molecule. It seemed to be of considerable interest to extend the measurements above 12,000 kg./cm.², both because of the light that such measurements might throw on the functioning of apparatus immersed in fluids up to 30,000, and because of the intrinsic interest from the point of view of theory of liquids. Furthermore, there is obviously an immediate application to the problems of geophysics.

Extension of the range of viscosity measurements to 30,000 demands a redesign of the apparatus. The former falling weight is now replaced by a swinging vane, which swings on a longitudinal axis back and forth between stops as the whole pressure apparatus is rotated about a longitudinal axis. With a given set-up viscosity is proportional to the time of swing. By suitable changes in the weight and the geometry of the vane it was possible to cover by actual measurement a range of 10⁷-fold in the absolute viscosity, a range a hundred-fold greater than with the former apparatus. The apparatus having been constructed, there were not many liquids which could be made the subject of measurement with it, since, as already stated, most liquids freeze at 30,000 kg./cm.² at room temperature. Some half a dozen liquids were measured over the entire range. Five of these, four alcohols and ethyl bromide properly, by thermodynamics, should have frozen but remained liquid by subcooling. The sixth liquid, pentane, is just on the verge of freezing under the extreme pressure. With these liquids the same effects were found in the extended pressure range that had been found previously to 12,000. The results are shown in figures 9 and 10. Eventually all the curves of logarithm of viscosity versus pressure turn upward. The pressure of upturn was highest for

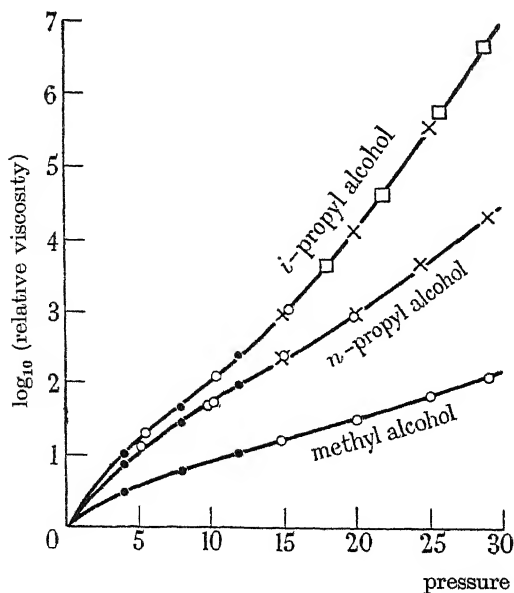


FIGURE 9. The viscosity of three liquids as a function of pressure.

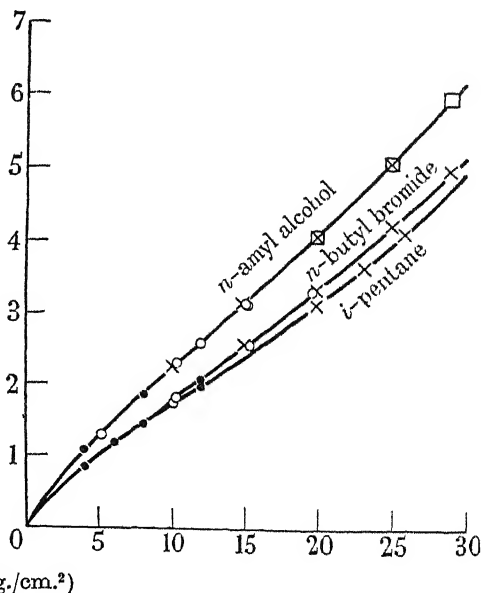


FIGURE 10. The viscosity of three liquids to 30,000 kg./cm.².

methyl alcohol, 20,000 kg./cm.². Of the six liquids measured methyl alcohol has the simplest molecular structure, so that the former generalization holds. The former generalization with regard to the connexion between the magnitude of the pressure effect and the complexity of the molecule also holds. These generalizations were emphasized by measurements on a new class of liquids, the dimethyl siloxanes, the molecules of which are highly complex. These liquids increase so rapidly in viscosity that measurements could not be carried beyond 12,000 kg./cm.², where the increase of viscosity already exceeded 10⁷-fold. At pressures of 30,000 such liquids must become as hard and stiff as the conventional glasses.

Finally, I shall briefly mention a different type of measurement which is still in progress so that only preliminary results have been yet obtained. These are measurements of the electrical resistance of metals up to 100,000 kg./cm.². There are obvious technical difficulties in such measurements. In the first place, pressure must be transmitted by a solid, since there are no liquids at these pressures at room temperature. The solid which transmits pressure must be an insulator, and there must be some method of minimizing or controlling the distortion produced by it in the metal under measurement. If one examines figure 7, in which is shown the apparatus for measuring volume changes up to 100,000 mounted within the 30,000 apparatus, I think it will be agreed that the difficulties are overwhelming against a straightforward extension of the same scheme of construction to resistance measurements, since this would demand getting leads through the walls of the 100,000 cylinder and all the accompanying manipulations. Some radically different method of approach is indicated.

The leading idea is the same as that which I had previously employed in realizing in steel apparatus shearing stresses up to the plastic flow limit combined with mean hydrostatic pressures of 50,000 kg./cm.². This pressure is much beyond the normal compressive strength of steel. The principle is that small surface areas on massive blocks may be stressed much beyond the stresses which normally produce fracture when the stresses are uniformly distributed, the highly loaded local parts receiving support from the surrounding unloaded regions. Simple examples of the operation of the principle are given by two crossed knife edges pressed against each other or by the cone of a Brinell hardness testing machine. Figure 11 illustrates the application to the present problem. Two carboloy blocks, 1 in. in diameter, reinforced with shrunk-on steel rings, confront each other between the platens of a hydraulic press. The two blocks are separated from each other by a layer of insulating material 0.010 in. thick and 0.500 in. in diameter. Embedded in the insulating material is the metal to be measured, in the form of a thin strip. The ends of the strip are brought through the insulator to make contact with the carboloy blocks, one end with one block and the other with the other. By attaching current and potential leads to the two carboloy blocks, one of which is insulated from the press, the resistance of the strip of metal embedded in the insulator may be measured. The carboloy blocks may be pushed together by the press with an intensity reaching up to 100,000 kg./cm.² over the area of contact with the insulator.

Details of the insulating region separating the carboloy blocks are shown in the lower part of figure 11. The outer edge of the insulator is a ring of pipe stone

(Catlinite) $\frac{1}{2}$ in. outside diameter and $\frac{13}{32}$ in. inside diameter. The rest of the insulating material is silver chloride, which is plastic and transmits pressure approximately hydrostatically. The function of the pipe-stone ring is to prevent, by friction against the carboly blocks, the lateral extrusion of the silver chloride. The crushing strength of the pipe stone is low enough, so that the load is distributed essentially uniformly over the entire area of contact. The deformation of the silver chloride and of the metal embedded in it is obviously not a true hydrostatic compression but is a one-dimensional shortening along the axis of the carboly blocks, since the change of dimensions of the carboly parallel to the face of the block is negligible. This means that the strain in the metal strip undergoing deformation is equivalent to a uniform hydrostatic compression on which has been superposed a uniform two-dimensional dilatation. This dilatation does not result in fracture of the metal because of the ductility imparted to metals by high pressure. It results that the distortion of the specimen is reproducible and specifiable so that the measured changes of resistance may be corrected to give the changes that would be found if the strain were a true hydrostatic compression.

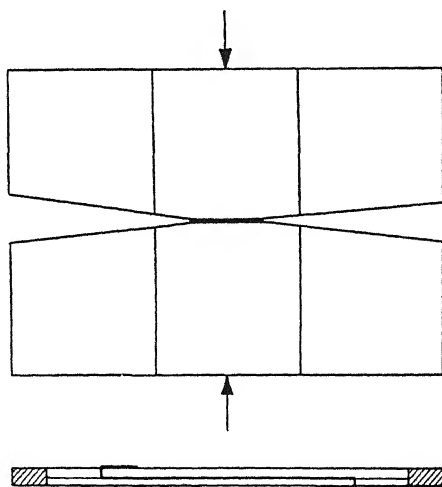


FIGURE 11. General scheme of the method for measuring the resistance of metals to 100,000 kg./cm.². In the lower part of the figure the metal is shown embedded in a plastic insulator (silver chloride) ready for mounting between the carboly blocks above, by which the pressure is applied.

Preliminary results have been obtained for a number of the more easily handled metals. In general, there is very marked deviation from linearity between pressure and change of resistance, as, indeed, there must be, since the metals measured are those whose resistance decreases under pressure. Although the curvature is marked, no instances have yet been found in which the resistance passes through a minimum and then increases. Examples of such behaviour had been previously found for the alkali metals, and it was surmised that other metals would behave similarly at higher pressures, but no such instances have yet been found. One abnormal metal has been measured, bismuth. The occurrence of the various polymorphic transitions

manifests itself by disturbances in the curve of resistance against pressure. The disturbances are, however, relatively slight, and it would appear that the precise lattice type is not an important factor in determining resistance under these conditions.

REFERENCES

- Adams, L. H. & Williamson, E. D. 1923 *J. Franklin Inst.* **195**, 493.
 Bardeen, J. 1938 *J. Chem. Phys.* **6**, 372.
 Lawson, A. W. & Ting-Yuan Tang 1949 *Phys. Rev.* **76**, 301.
 Sternheimer, R. M. 1949 *Phys. Rev.* **75**, 888.

Kinetic salt effects on the inversion of sucrose

By E. A. GUGGENHEIM, F.R.S., *Department of Chemistry, Reading University*
 AND L. A. WISEMAN, *Explosives Research and Development Establishment,*
Waltham Abbey

(Received 27 April 1950)

The rate of inversion of sucrose by strong acids has been measured at 24.7° C by a method which eliminated the mutarotation lag and the end-point uncertainty. The rate was proved to be strictly first order for at least 54 hr. and the first-order rate constant was determined with an internal consistency better than 1 %.

In 0.1 M hydrochloric acid the first-order rate constant k_1 is linear in S , the number of grams of sucrose in a litre, according to the formula

$$k_1 \times 10^4 / \text{min}^{-1} = 7.13 + 0.97 \times 10^{-2} S.$$

For sucrose at 30 g./l. hydrolysed by single strong acids at molarities up to 0.2, the first-order rate constant k_1 is related to the molarity c of strong acids by the formula

$$k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{B_j},$$

where B_j is a constant determined by the anion as follows:

$$\text{Cl}^-, 0.28; \text{Br}^-, 0.35; \text{ClO}_4^-, 0.38; \text{NO}_3^-, 0.30.$$

In similar experiments with mixed strong acids or a strong acid with the addition of a neutral uni-univalent salt, the experimental results can be expressed by the formula

$$k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c_H \times 10^{\Sigma B_j c_j},$$

where c_H denotes the molarity of strong acids and c_j the molarity of each anion. Each B_j has the same value as in the formula for single acids. This formula is in accordance with Bronsted's principle of specific interaction.

In experiments on solutions containing bi-univalent and tri-univalent salts there is a further small negative salt effect of the multivalent cations increasing with their valency. This effect indicates that the principle of specific interaction becomes detectably inaccurate at these higher ionic strengths.

Our results are compared with those of other workers. On the whole there is good agreement within the experimental accuracy of the various data. Certain discrepancies in absolute values of the rates are probably attributable to uncertainties of 0.1° C or less in the temperature.

1. INTRODUCTION

The object of the research here described was to determine as accurately as possible in dilute aqueous solutions the magnitude of the kinetic salt effects on the inversion of sucrose catalysed by hydrogen ions. The experimental work described was carried

out in the Chemistry Department of University College, London, during the years 1936–8. It was then interrupted. At that time we hoped that the work might be resumed at some later date, but we now realize that for reasons beyond our control the opportunity for this will not arise. Although our original programme is thus incomplete, we believe that the results obtained are of sufficient interest to merit description.

2. GENERAL THEORY

The foundation of the theory of primary kinetic salt effects was laid by Brönsted (1922*a*). All subsequent progress has been built on this foundation. Reviews of the experimental support of the theory have been given by Livingston (1930) and by La Mer (1932). Brönsted's fundamental equation for the bimolecular rate constant k_2 of a reaction involving a molecule A and a molecule B , and possibly also the solvent, is

$$k_2 = k_2^0 f_A f_B / f_X, \quad (2.1)$$

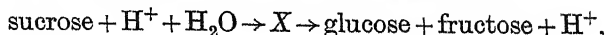
where k_2^0 is characteristic of the reaction, the solvent and the temperature but is independent of the concentrations of the reactants or any other substances present in dilute solutions; f_A , f_B , f_X denote the activity coefficients of the reactant A , the reactant B and the reacting complex X formed by A and B in collision.

To make quantitative use of (2.1) we require formulae expressing the activity coefficients in terms of the composition of the solution. For this purpose it is convenient to use formulae proposed by one of us (Guggenheim 1935, 1936), wherein the formulae of Debye & Hückel (1923) have been adapted so as to embrace Brönsted's (1922*b*) principle of specific interaction of ions. We accordingly write for the activity coefficient f_i of an ion of type i having a valency z_i

$$\log_{10} f_i = -Az_i^2 \frac{I^{\frac{1}{2}}}{1 + I^{\frac{1}{2}}} + \sum_j B_{ij} c_j, \quad (2.2)$$

where I denotes ionic strength, c_j denotes the concentration of ions of type j , A depends only on the solvent and the temperature; $B_{ij} = B_{ji}$ depends on the solvent, the temperature and the nature of the ions i and j . In accordance with the principle of specific interaction, the summation \sum_j extends only over ions of opposite charge to the ion i . Formula (2.2) has been shown (Guggenheim 1935, 1936) to be in good agreement with the best experimental data in aqueous solutions at ionic strengths up to 0.1. It has further proved useful in correlating experimental data on acid base indicator equilibria (Guggenheim & Schindler 1934).

For the inversion of sucrose catalysed by hydrogen ions



formula (2.1) becomes

$$k_2 = k_2^0 f_S f_{\text{H}} / f_X, \quad (2.3)$$

where S denotes sucrose. The reacting complex X has a single positive charge the same as the hydrogen ion. Consequently, when we use (2.2), the terms proportional to Az_i^2 cancel and we are left with

$$k_2 = k_2^0 \times 10^{\sum -B_j c_j}, \quad (2.4)$$

where the summation Σ_- extends over all anions, but not cations. In the simple case when the solution contains no electrolyte except a single strong acid such as HCl, formula (2.4) reduces to

$$k_2 = k_2^0 \times 10^{Bc}, \quad (2.5)$$

where c now denotes the concentration of the strong acid. k_2^0 will depend on the temperature and possibly on the sucrose concentration if this is high enough to affect appreciably the physical properties of the medium. Formulae of the type (2.5) have often been used empirically (see, for example, Schmid & Olsen 1926). Brønsted (1926) himself derived formulae equivalent to (2.4) and (2.5), starting from a formula differing from (2.2) by the omission of the factor $1 + I^{\frac{1}{2}}$ in the denominator of the first term. Such a derivation is valid only for ionic strengths less than 0.01, whereas our derivation is sound for ionic strengths up to about 0.1.

When we describe our experimental results we shall compare them with these theoretical formulae. We shall also use these formulae in §11, where we discuss experimental results of other workers.

3. MUTAROTATION LAG

It is well known that the hydrolysis of sucrose into glucose and fructose is followed by the mutarotation of both products at rates greater by a factor of about 10^2 than that of the hydrolysis. Thus the optical rotation at any moment will differ by a small but not constant amount from the value it would have if mutarotation were instantaneous. This is sometimes called the 'mutarotation lag'.

In order to obtain the best conditions for the theoretical interpretation of the kinetic salt effects it was desirable to work at low concentrations. Under these conditions the salt effects to be measured are small, and so the highest attainable accuracy was desirable. A rough estimate indicates that the mutarotation lag, if not eliminated, could introduce an error in the rate constant less than 1 % but greater than 0.1 %. We decided initially to aim at an accuracy better than 1 %, and it therefore seemed desirable to eliminate the mutarotation lag. We eventually came to the conclusion that our measurements of the rate constant were in fact accurate to 1 % but no better. As far as accuracy is concerned it would therefore have mattered little if we had not troubled to eliminate the lag.

There is, however, another and more important reason for eliminating the mutarotation lag. It has been alleged (Pennycuik 1926; but see Scatchard 1926) that the inversion of sucrose is not exactly a first-order reaction. We wished to resolve this question unambiguously, and for this purpose it seemed essential to eliminate the mutarotation lag. We achieved this by the following procedure.

Samples of the reacting mixture were neutralized with a slight excess of alkali before measurement of the optical rotatory power. Neutralization of the acid completely stops the hydrolysis of the sucrose and the further addition of a minute excess of alkali completes mutarotation practically instantaneously.

A somewhat similar technique was used by Fales & Morell (1922), but these authors added only enough alkali to make 'neutral to litmus', which means a pH somewhere between 6 and 8. If the pH were nearer 6 than 8 the mutarotation would be far from instantaneous.

4. MATERIALS USED

The sucrose was an Analar product obtained in large batches, so that samples from the same batch could be used in as many runs as possible. The several batches were intercompared and no systematic difference was detected. The sucrose was obtained as large crystals which were ground into a powder and then dried in a desiccator for at least twelve hours. Tests for moisture and ash content gave results well within the Analar specifications.

The hydrochloric acid solutions were made up by weight from the constant-boiling acid. The hydrobromic acid used was Analar constant-boiling acid which was redistilled to remove bromine, the absence of which was confirmed by testing with potassium iodide and starch. The perchloric acid was an Analar 20 % solution and the nitric acid was also Analar. All the acid solutions were standardized by titration against Analar sodium carbonate.

The salts used, except the lanthanum chloride supplied by Hopkins and Williams, were Analar products. The calcium chloride and magnesium chloride were made by dissolving calcium carbonate and magnesium oxide respectively in the required amount of hydrochloric acid. All the reaction solutions containing salts as well as acid were standardized against Analar sodium carbonate.

5. EXPERIMENTAL PROCEDURE

The reaction solutions were made up in 1 l. graduated flasks which had been calibrated. The sucrose was dissolved in a small quantity of distilled water, which had been brought to 25°C, the requisite amount of acid (by volume) and salt (by weight) were then added and then the whole was made up to 1 l. of solution. This was left in the thermostat (set at about 25°C and constant to better than 0.1°C) for half an hour, so that the volume was definitely 1 l. at the thermostat's temperature, and then the solution was transferred to a Pyrex Erlenmeyer flask, fitted with a ground-glass stopper. The homogeneity of the solution was assured by frequent shaking. According to the only standard thermometer available the temperature of the thermostat was 24.68°C, but as we have no record of when the thermometer was standardized and have no means of cross-checking, we consider it prudent to admit the possibility of an uncertainty of say $\pm 0.05^\circ\text{C}$. Whatever this temperature may have been, it was maintained constant to better than 0.01°C. After the solution had been in the thermostat for 1½ to 2 hr. the first measurement was made.

For each measurement 40 ml. of the reaction solution was removed, neutralized with 2N alkali at a known time, and then the neutralized solution was made up to 50 ml. with distilled water, at 25°C. The solution was then transferred to a 40 cm. polarimeter tube, jacketed at 25°C and its optical rotation measured on a Hilger polarimeter, reading to 0.01°. A high-intensity sodium vapour lamp supplied by the General Electric Company and working off the a.c. mains was used as the source of illumination. In most of the experiments sodium or potassium hydroxides were used for neutralization. For the experiment with 0.20 M perchloric acid we had to use sodium hydroxide, since with potassium hydroxide the sparingly soluble potassium perchlorate would be formed. For the experiments with the alkaline earth

chlorides, we used carbonate-free sodium hydroxide (i.e. with a small amount of barium hydroxide to remove the carbonate) to prevent the formation of the insoluble alkaline earth carbonates.

The quantity of alkali used for neutralization was in every case a slight excess so that the solution in the polarimeter tube was between 0.002N and 0.004N in free alkali. This slight excess of alkali completes the mutarotation in a minute or less.

This technique raises the following questions:

- (1) Does the value of the optical rotation depend on the time after neutralization?
- (2) Does the value of the optical rotation depend on the concentration of free alkali?
- (3) Does the value obtained for the rate constant depend on the nature of the alkali?

These questions were satisfactorily answered as follows.

We measured the optical rotation of neutralized samples of the reaction solution at times from 20 min. up to 24 hr. after neutralization. It was found that the rotation was constant over this period of time, except when the samples were taken from reaction solutions which were 48 hr. old or more, in which case there was a change of about 0.05° in 24 hr. As the measurements were never taken more than three hours after neutralization, and generally within one hour, and the polarimeter could be read only to 0.01° , we decided that we could neglect this slight change.

The salts present in the sugar solution have an appreciable effect on the rotation. In neutralizing with alkali, we could not add exactly the same excess of alkali each time, the maximum excess being about 0.004N. We verified, however, that the effect of free alkali at these low concentrations was negligible by measuring the effect of 0.02N free alkali.

As alkalis we used potassium hydroxide or sodium hydroxide, or in a few experiments a mixture of sodium and barium hydroxides. A few pairs of experiments were done with the same sucrose and acid concentrations (as acids we used hydrochloric and hydrobromic acids), but using in the one case sodium hydroxide and in the other potassium hydroxide as the neutralizing alkali. The rate constants so obtained agreed within 1 %.

6. CALCULATIONS OF RATE CONSTANT

There are several methods for calculating a first-order rate constant k_1 from readings r at times t of some physical property such as the optical rotation. The most straightforward method relies on the formula

$$\log(r - r_\infty) = a - k_1 t, \quad (6.1)$$

where r_∞ is the *observed* value of r as $t \rightarrow \infty$, while the constants a and k_1 are adjusted to give the best fit. This method has the advantage of convenience, but it is quite unsuitable for obtaining the best value for the rate constant k_1 because a random error δ in r_∞ is mathematically equivalent to a systematic error $-\delta$ in all the readings r . Hence for accuracy this method requires that r_∞ should be known with much greater accuracy than any of the other readings r , whereas r_∞ is in fact usually known with considerably less accuracy than the other r 's. There are at least two known ways of avoiding this inaccuracy due to uncertainty in r_∞ .

In the first method (Guggenheim 1926), which for brevity we shall call the G method, each reading r at time t is paired with a reading r' at a time $t + \tau$, where τ is a suitably chosen constant interval. We have then the formula

$$\log(r' - r) = a' - k_1 t. \quad (6.2)$$

It is thus only necessary to plot $\log(r' - r)$ against t and measure the slope of the best straight line. The accuracy of the method can, if desired, be checked by using several alternative intervals τ . Even so the method is open to the criticism that the drawing of the best straight line may be subject to personal bias. Nor can the best straight line be determined by direct application of the method of least squares to (6.2) because the least squares treatment if used at all should be applied to $(r' - r)$, not to $\log(r' - r)$. Admittedly the method might reasonably be applied directly to $\log(r' - r)$ provided the points are suitably weighted by some factor such as $(r' - r)$ or $(r' - r)^2$, but the selection of such a weighting factor would again depend on personal choice.

In the second method, due to Bond (1935), which for brevity we shall call the B method, the least squares treatment is applied, as it should be, to the readings r . That is to say, the three constants k_1 , A , B are adjusted so as to minimize

$$\sum_i (r_i - Ae^{-k_1 t} - B)^2.$$

This procedure gives the best value of k_1 and at the same time leads to the best value of r_∞ , namely B . For details of this method we refer the reader to Bond's book. It is unquestionably the right method for obtaining the most accurate value for k_1 . The only possible objection is that it is laborious, each calculation of k_1 taking nearly two hours.

We have compared the G method with the B method and found that they agreed to within better than 1 %. Nevertheless, in spite of the extra labour, we used the B method in all cases, and the quoted values of k_1 were thus obtained. A further advantage of the B method is that it leads to an estimate of the probable accuracy in k_1 . This was found to be better than $\frac{1}{2}$ %.

Having obtained the best values of k_1 , A , B by the B method we calculated values of r for each time t at which a reading was taken. To illustrate the accuracy obtained we give in tables 1.1, 1.2 and 1.3 comparisons of observed and calculated values of r for three experiments. It will be seen that there is agreement within $\pm 0.01^\circ$, which is the accuracy with which the polarimeter could be read.

It is clear from these tables and the other similar tables that the reaction is accurately first order at least over the first 54 hr. In a few cases we compared calculated with observed values of r after longer times t . We found that agreement to within ± 0.01 persisted for three whole days, but after a fourth day the discrepancy was sometimes 0.02 or 0.03, and in one case was 0.06. This behaviour suggests the occurrence of uncontrolled changes due to bacteria or moulds, and we did not attempt to investigate it further. It does, however, strengthen our profound mistrust of the accuracy of values of k_1 based on an allegedly observed value of r_∞ .

The values of r_∞ calculated by the B method varied from one experiment to another within ± 0.05 , which is about $\frac{1}{2}$ % of the total change in r . No significance can be attached to the absolute values of r . The polarimeter scale was set so that the final

TABLE 1.1. 0.05M HCl

r calculated from $r = r_{\infty} - (r_{\infty} - r_0) e^{-k_1 t}$; $r_0 = 8.11$, $r_{\infty} = 0.20$, $k_1 = 0.0215 \text{ hr.}^{-1} = 3.58 \times 10^{-4} \text{ min.}^{-1}$.

t (hr.)	r (exp.)	r (calc.)	$r(\text{exp.}) - r(\text{calc.})$
0.00	8.12	8.11	+0.01
2.00	7.77	7.78	-0.01
4.00	7.46	7.46	0.00
6.00	7.16	7.15	+0.01
22.00	5.13	5.13	0.00
24.00	4.91	4.92	-0.01
26.00	4.72	4.72	0.00
28.00	4.52	4.53	-0.01
30.00	4.35	4.35	0.00
46.00	3.14	3.14	0.00
48.00	3.01	3.02	-0.01
50.00	2.91	2.90	+0.01
54.00	2.66	2.68	-0.02

TABLE 1.2. 0.10M HCl

r calculated from $r = r_{\infty} - (r_{\infty} - r_0) e^{-k_1 t}$; $r_0 = 8.17$, $r_{\infty} = 0.24$, $k_1 = 0.0445 \text{ hr.}^{-1} = 7.42 \times 10^{-4} \text{ min.}^{-1}$.

t (hr.)	r (exp.)	r (calc.)	$r(\text{exp.}) - r(\text{calc.})$
0.00	8.16	8.17	-0.01
2.00	7.49	7.49	0.00
4.00	6.89	6.88	+0.01
6.50	6.17	6.18	-0.01
22.00	3.20	3.22	-0.02
24.00	2.96	2.97	-0.01
26.00	2.73	2.73	0.00
28.00	2.54	2.52	+0.02
30.50	2.28	2.28	0.00
46.00	1.26	1.26	0.00
48.00	1.18	1.18	0.00
50.00	1.10	1.10	0.00
54.50	0.94	0.94	0.00

TABLE 1.3. 0.20M HCl

r calculated from $r = r_{\infty} - (r_{\infty} - r_0) e^{-k_1 t}$; $r_0 = 7.10$, $r_{\infty} = 0.255$, $k_1 = 0.0947 \text{ hr.}^{-1} = 15.78 \times 10^{-4} \text{ min.}^{-1}$.

t (hr.)	r (exp.)	r (calc.)	$r(\text{exp.}) - r(\text{calc.})$
0.00	7.11	7.10	+0.01
2.00	5.92	5.92	0.00
4.00	4.94	4.94	0.00
6.00	4.13	4.13	0.00
22.00	1.13	1.11	+0.02
24.00	0.94	0.96	-0.02
26.00	0.84	0.84	0.00
28.00	0.74	0.74	0.00
30.00	0.66	0.65	+0.01
46.00	0.35	0.34	+0.01
48.00	0.33	0.33	0.00
50.00	0.32	0.32	0.00
54.00	0.27	0.30	-0.03

reading had a small positive value. The position of the scale was never intentionally changed. Any small accidental shift between experiments would of course affect the value of r_∞ but not of k_1 .

7. TEMPERATURE COEFFICIENT

It was not part of our programme to study the temperature dependence of the rate constant. This has been studied by numerous other workers, and with particular thoroughness by Leininger & Kilpatrick (1938). Their results had, however, not yet been published at the time of our experiments. When we found out that our nominal temperature of 25°C was actually 24.68°C according to our only 'standard' thermometer, we decided to make a few measurements at 30°C to enable us, if desired, to correct our rate constants to 25°C. We made measurements with both hydrochloric and perchloric acids at molarities 0.05, 0.10, 0.15, 0.20. In both cases we found that as the molarity increases from 0.05 to 0.20 the temperature coefficient decreases by something between 2 and 3 %, which we believe to be outside our experimental error, although our data are insufficient to establish this effect beyond doubt. However, we can say with confidence that over the whole range of concentrations

$$\frac{k \text{ at } 30^\circ \text{C}}{k \text{ at } 24.68^\circ \text{C}} = 2.16 \pm 0.03,$$

and this is sufficient to deduce by interpolation that

$$\frac{k \text{ at } 25^\circ \text{C}}{k \text{ at } 24.68^\circ \text{C}} = 1.048.$$

If, then, the temperature of most of our experiments was truly 24.68°C as read on our only standardized thermometer, the observed values can be corrected to 25°C by multiplication by 1.048.

8. DEPENDENCE ON SUCROSE CONCENTRATION

Most of our measurements were made on solutions containing 30 g. of sucrose in 1 l. of solution at 25°C, this being about the lowest concentration at which we had confidence in the accuracy obtainable. We however made one series of measurements to study the effect of varying the sucrose concentration. These experiments were made using 0.1 M hydrochloric acid. The results are given in table 2, where the measured values of the first-order rate constant k_1 are compared with values calculated from the empirical formula

$$k_1 \times 10^4 / \text{min.}^{-1} = 7.13 + 0.97 \times 10^{-2} S, \quad (8.1)$$

where S denotes the number of grams of sucrose in a litre. It will be seen that the agreement is better than 1 % and usually better than $\frac{1}{2}$ %.

Since a molar solution of sucrose contains 342 g. in a litre, we see that the medium effect of 0.1 M sucrose is to increase the rate by about 5 % which, as we shall see, is less than but comparable with the salt effects at 0.1 M.

TABLE 2. 0.1 M HCl

S denotes sucrose concentration in g./l. k_1 denotes first-order rate constant and is calculated from

$$k_1 \times 10^4 / \text{min.}^{-1} = 7.13 + 0.97 \times 10^{-2} S$$

S (g./l.)	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
30	7.42	7.42
30	7.50	7.42
30	7.43	7.42
30	7.47	7.42
35	7.47	7.47
40	7.52	7.52
60	7.73	7.71
80	7.88	7.91
100	8.08	8.10

9. MEASUREMENTS WITH SINGLE ACIDS

We made measurements on solutions containing 30 g. of sucrose per litre at 25°C using four different strong acids, namely, hydrochloric, hydrobromic, perchloric and nitric. In each case the measurements extended over the range of molarities from 0.05 to 0.2. According to the theory outlined in §2 the bimolecular constant k_2 , obtained by dividing the observed first-order constant k_1 by the molarity c of the strong acid, should obey a relation of the form (2.5). This was verified in all four cases with an accuracy of 1 % or better. The experimental results can be represented by the following formulae:

$$\text{HCl: } k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.28c},$$

$$\text{HBr: } k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.35c},$$

$$\text{HClO}_4: k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.38c},$$

$$\text{HNO}_3: k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.30c},$$

where in each case c denotes the molarity of strong acid.

The comparison between observed and calculated values of k_1 is given in tables 3.1, 3.2, 3.3 and 3.4. Out of twenty-six experiments there are three in which the

TABLE 3.1

c denotes molar concentration of HCl. k_1 denotes first-order constant and is calculated from $k_1 / \text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.28c}$.

c	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
0.05	3.58	3.59
0.05	3.55	3.59
0.05	3.58	3.59
0.075	5.40*	5.47
0.075	5.47	5.47
0.10	7.42	7.42
0.10	7.50*	7.42
0.10	7.43	7.42
0.10	7.47	7.42
0.15	11.48	11.48
0.20	15.77	15.81
0.20	15.78	15.81
0.20	15.73	15.81

TABLE 3.2

c denotes molar concentration of HBr. k_1 denotes first-order constant and is calculated from $k_1/\text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.35c}$.

c	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
0.05	3.63	3.62
0.05	3.60	3.62
0.10	7.65*	7.53
0.20	16.41	16.33
0.20	16.22	16.33

TABLE 3.3

c denotes molar concentration of HClO_4 . k_1 denotes first-order constant and is calculated from $k_1/\text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.38c}$.

c	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
0.05	3.64	3.63
0.10	7.55	7.59
0.15	11.97	11.88
0.20	16.60	16.55

TABLE 3.4

c denotes molar concentration of HNO_3 . k_1 denotes first-order rate constant and is calculated from $k_1/\text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{0.30c}$.

c	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
0.05	3.63	3.60
0.10	7.52	7.45
0.15	11.63	11.56
0.20	15.82	15.96

discrepancy between observed and calculated values exceeds 1%. These are indicated by an asterisk. In the case of nitric acid only there appears to be a slight trend in the discrepancies; the experimental values are higher than the calculated at the lower molarities, but lower than the calculated at the higher molarities. Since, however, none of these deviations exceed 1% this appearance of trend may be fortuitous. It is unfortunate that we were unable to settle this question by further experiments.

10. MEASUREMENTS WITH MIXED ACIDS OR WITH ADDED SALTS

In order to test the applicability of the principle of specific interaction of ions we made experiments with mixtures of two strong acids and experiments in which a salt was present as well as the strong acid. All these experiments were made on solutions containing 30 g. of sucrose per litre. The results are given in table 4, where the experimental values of k_1 divided by the molarity c_H of the strong acid are compared with the calculated values.

The calculated values were obtained in accordance with (2.4) from the formula

$$k_1/\text{min.}^{-1} = 6.95 \times 10^{-3} c \times 10^{2-B_2 c_j},$$

where c_j denotes the molar concentration of each anion and the coefficients B_j have the values already determined from the experiments with single strong acids alone, namely,

$$\text{Cl}^-: B_j = 0.28 \text{ l./mole}$$

$$\text{Br}^-: B_j = 0.35 \text{ l./mole}$$

$$\text{ClO}_4^-: B_j = 0.38 \text{ l./mole}$$

$$\text{NO}_3^-: B_j = 0.30 \text{ l./mole}$$

The table also includes a measurement made with 0.1 M hydrochloric acid containing 0.1 M potassium iodide. In this case we found that after 3 days the solution contained free iodine equivalent to 1 % of the potassium iodide. There was no evidence of any chemical effect of this iodine, since the reaction remained accurately first order and the extrapolated end-point did not differ significantly from its usual value. Since the acid concentration was decreased by only 1 % in 3 days the rate constant should not be affected by more than 1 %. We cannot have the same confidence in this result as in the measurements without any iodide, but we quote it for what it is worth. If it is reliable, then for iodide we have

$$\text{I}^-: B_j = 0.40 \text{ l./mole},$$

so that the B_j value for bromide is nearly the mean of those for chloride and iodide, which is at least reasonable.

In table 4 the cases where the discrepancy between observed and calculated values exceeds 1 % are indicated by an asterisk. It will be noticed that among the solutions containing only univalent ions there are just two for which the discrepancy slightly exceeds 1 %. On the other hand, the discrepancies exceed 1 % for all the four solutions containing multivalent cations. The deviations lie between 1 % and 2 % for the solutions containing bivalent cations and it is over 3 % for the solution containing the trivalent lanthanum ion. These deviations are well outside our usual experimental error. They probably signify that the formulae given in §2 are breaking down

TABLE 4

k_1 calculated on assumption that the anions contribute additively to $\log k_1/c_{\text{H}^+}$ and that there is no specific effect of cations.

composition of solution in mole/l.	$k_1 \times 10^4$ (min. ⁻¹)	$10^3 k_1/c_{\text{H}}$ (exp.)	$10^3 k_1/c_{\text{H}}$ (calc.)
0.10 HCl + 0.10 HBr	16.22	8.11	8.04
0.10 HCl + 0.10 HClO ₄	16.35	8.17	8.09
0.10 HBr + 0.10 HClO ₄	16.45	8.22	8.22
0.10 HCl + 0.10 NaCl	7.89	7.89	7.91
0.10 HCl + 0.10 KCl	7.93	7.93	7.91
0.10 HCl + 0.10 KBr	7.95	7.95*	8.04
0.10 HCl + 0.10 KI	8.13	8.13	—
0.10 HClO ₄ + 0.10 NaCl	8.05	8.05	8.09
0.05 HNO ₃ + 0.10 KNO ₃	3.85	7.70	7.71
0.10 HNO ₃ + 0.10 KNO ₃	7.90	7.90*	7.98
0.107 HCl + 0.0465 MgCl ₂	8.38	7.83*	7.91
0.10 HCl + 0.05 CaCl ₂	7.80	7.80*	7.91
0.10 HCl + 0.05 BaCl ₂	7.75	7.75*	7.91
0.10 HCl + 0.0333 LaCl ₃	7.67	7.67*	7.91

owing to the ionic strength being too high. This is not altogether surprising since the ionic strength of the solution containing lanthanum chloride is 0.3, whereas the principle of specific interaction in the form used in §2 is not claimed to be accurate at ionic strengths appreciably exceeding 0.1.

It is interesting to note that the deviations due to the presence of multivalent cations have the sign which might reasonably be expected. Since the effect of anions is to increase the rate, it is reasonable to suppose that cations should decrease the rate and the more so the higher the valency.

11. COMPARISON WITH OTHER WORK

The number of papers concerned with the inversion of sucrose is so great that we could not refer to them all, even if we wished to do so. We shall rather confine ourselves to mentioning two classes of papers: (a) those in which the experiments described have a direct bearing on kinetic salt effects, and (b) those describing experiments at 25°C of sufficient accuracy for profitable comparison with our own experimental results.

The famous paper of Arrhenius (1889) describes a large number of measurements at 39.5°C and at 52.5°C on solutions containing 100 g. of sucrose per litre. Whereas addition of 1 M non-electrolytes produced increases in rate varying between 0% and 10%, addition of 0.4 M chlorides of either univalent or bivalent metals produced an increase of about 25%. Our results at 25°C if extrapolated indicate an increase of about 30% on addition of 0.4 M chloride.

Palmaer (1897) measured the rate of inversion of sucrose 100 g./l. at 48.18°C by hydrochloric acid at molarities varied between 0.001 and 0.1. According to these measurements k_1/c increases by 5.4% as the molarity of acid is increased to 0.1 from much smaller values. We find at 25°C an increase of 6.7%.

Euler (1900) published some measurements made in his laboratory by Miss Molin. These measurements were made at 28°C at several concentrations of sucrose with hydrochloric acid and potassium chloride at various molarities. As these measurements are closely related to ours and appear to be of higher internal consistency than many later measurements, we give a full analysis of them in table 5. We find that the combined results of the fifteen experiments can be expressed by the formula

$$k_1/\text{min.}^{-1} = 1.06 \times 10^{-2} c_{\text{H}} (1 + 10^{-3} S) \times 10^{0.20c},$$

where S denotes concentration of sucrose in g./l., c_{H} denotes molar concentration of hydrochloric acid and c denotes total molar concentration of chloride. It will be observed that according to this formula the dependence of k_1 on each of the variables c_{H} , S and c is of the same form as in our own work. The agreement between experimental and calculated values is throughout within 2% except for one experiment indicated by an asterisk. Comparing our own work with these experiments done about fifty years ago, all we can claim is to have reduced the random errors by a factor 2.

Fales & Morell (1922) measured the rate of inversion of 0.3 M sucrose, i.e. 102.6 g./l. at 35°C by hydrochloric acid of various concentrations. Their results show that the bimolecular rate constant increases with the concentration. Unfortunately, the method used for computing the rate constant is so naïve that the published values

have an uncertainty of several units per cent. A comparison of their values with those of Leininger & Kilpatrick (1938), mentioned below, confirms that their published values are inaccurate.

Bowe (1927) measured the inversion of sucrose 130 g./l. at 25°C by 0.1 M hydrochloric acid and studied the effect of added salts but only at molarities of 1 and greater. For 0.1 M hydrochloric acid without addition of salt he found

$$k_1 = 0.0514 \text{ hr.}^{-1} = 8.57 \times 10^{-4} \text{ min.}^{-1}.$$

Our value 0.1 M hydrochloric acid extrapolated to 130 g./l. is $k_1 = 8.39 \times 10^{-4} \text{ min.}^{-1}$ at 24.68°C, corresponding to $k_1 = 8.79 \times 10^{-4} \text{ min.}^{-1}$ at 25°C. The 3 % discrepancy between our value and Bowe's corresponds to a difference of 0.15°C between the temperatures.

Kautz & Robinson (1928) measured the inversion of sucrose 200 g./l. at 25°C by 0.1 M hydrochloric acid and studied the effect of adding various chlorides. According to their results $\log k$ is approximately linear in the concentration of chloride but there is considerable scatter. This may be partly due to the use of a crude and naive method of computing k_1 . The units of k_1 are not stated, but we suppose the quoted values to be $k_1/\log_e 10 \text{ min.}^{-1}$. According to this interpretation their value for 0.1 M hydrochloric acid without added salt is

$$2.303 \times 4.30 \times 10^{-4} \text{ min.}^{-1} = 9.90 \times 10^{-4} \text{ min.}^{-1}.$$

Our value for 0.1 M hydrochloric acid extrapolated to a sucrose concentration of 200 g./l. is $9.07 \times 10^{-4} \text{ min.}^{-1}$ at 24.68°C corresponding to $9.50 \times 10^{-4} \text{ min.}^{-1}$ at 25°C. As the internal consistency of our data is appreciably better than that of these authors, we attribute the 4 % discrepancy to their experimental error.

Kappanna & Shrikhande (1931) published experiments which are especially interesting because the rate of hydrolysis was determined by direct estimation of the glucose formed with Fehling's solution. In this method there is no mutarotation lag. The accuracy claimed is 2 %. The results of their experiments can be summarized as follows:

(a) In a series of experiments with 0.1 M hydrochloric acid at 35°C the dependence of k_1 on S , the number of grams of sucrose per litre, is given by

$$k_1 \times 10^4 / \text{min.}^{-1} = 33.65 + 1.42 \times 10^{-2} S.$$

This formula is valid over a range of sucrose concentrations extending from 1 to 100 g./l. and the agreement with the experimental values is better than 1 % at sucrose concentrations exceeding 5 g./l.; as might be expected from the method of analysis, the agreement is not so good at lower concentrations.

(b) A comparison of measurements with 0.1 M hydrochloric acid at two temperatures shows that

$$\frac{k \text{ at } 45^\circ \text{C}}{k \text{ at } 35^\circ \text{C}} = 3.50_5 \pm 0.005,$$

independent of the sucrose concentration. This ratio is somewhat smaller than the value 3.66 obtained by a slight extrapolation of Leininger & Kilpatrick's (1938) measurements.

(c) Measurements on sucrose 10 g./l. at 35° C with varying concentrations of hydrochloric acid can be expressed by either of the formulae

$$k_1/\text{min.}^{-1} = 32.0 \times 10^{-3}c \times 10^{0.174c},$$

$$k_1/\text{min.}^{-1} = (32.0 + 14c) \times 10^3c.$$

The former formula represents the experimental data within an accuracy better than 1 % for concentrations c extending from 0.001 M to 0.6 M. Surprisingly the latter formula agrees with the experimental data within 1 % over the wider range 0.001 M to 1 M. This behaviour is the opposite of that found by Leininger & Kilpatrick (1938).

(d) Measurements were made on sucrose 10 g./l. at 35° C with 0.1 M hydrochloric acid and addition of chloride of potassium, sodium or lithium at various concentrations. The results for total chloride molarities not exceeding 0.3 can be expressed by the formula

$$k_1/\text{min.}^{-1} = 32.0 \times 10^{-4} \times 10^{0.174c_{\text{Cl}}},$$

where c_{Cl} denotes total chloride molarity. The agreement between measured and calculated values of k_1 is, except in one experiment, better than 2 %.

Floyd (1931) measured the inversion of 0.1 M sucrose, i.e. 34.2 g./l. by 0.1 M hydrochloric acid and studied the effect of addition of sodium chloride or barium chloride mostly at molarities of 1 or greater. Most of the measurements were made at 30° C with some at 25° C. For 0.1 M hydrochloric acid at 25° C without any salt added he obtained $k_1 = 7.93_5 (\pm 0.03) \times 10^{-4} \text{ min.}^{-1}$. Our value for the same sucrose concentration is $k_1 = 7.47 \times 10^{-4} \text{ min.}^{-1}$ at 24.68° C corresponding to $k_1 = 7.83 \times 10^{-4} \text{ min.}^{-1}$ at 25° C. The discrepancy of 1.2 % between our value and Floyd's corresponds to a temperature difference of 0.08° C. The effect of adding 0.1 M sodium chloride to the 0.1 M hydrochloric acid at 30° C increased the rate by a factor 1.03₅ directly measured or 1.05₂ when the data for various concentrations of sodium chloride are smoothed. Our corresponding factor at 25° C is 1.06₅. For the temperature coefficient Floyd found with 0.1 M hydrochloric acid

$$\frac{k \text{ at } 30^\circ \text{ C}}{k \text{ at } 25^\circ \text{ C}} = 2.05_7,$$

whereas we found

$$\frac{k \text{ at } 30^\circ \text{ C}}{k \text{ at } 24.68^\circ \text{ C}} = 2.19,$$

corresponding to

$$\frac{k \text{ at } 30^\circ \text{ C}}{k \text{ at } 25^\circ \text{ C}} = 2.08.$$

Hitchcock & Dougan (1935) studied the hydrolysis of sucrose by hydrochloric acid at 25° C using a dilatometric method. From these measurements we find for 0.1 M hydrochloric acid by a slight extrapolation:

$$\begin{aligned} \text{sucrose 20 g./l.: } k_1 &= 2.303 \times 3.34_5 \times 10^{-4} \text{ min.}^{-1} \\ &= 7.70 \times 10^{-4} \text{ min.}^{-1}, \end{aligned}$$

$$\begin{aligned} \text{sucrose 50 g./l.: } k_1 &= 2.303 \times 3.46_7 \times 10^{-4} \text{ min.}^{-1} \\ &= 7.98 \times 10^{-4} \text{ min.}^{-1}, \end{aligned}$$

whence by interpolation

$$\text{sucrose } 30 \text{ g./l.: } k_1 = 7.79 \times 10^{-4} \text{ min.}^{-1}.$$

Our measured value is $k_1 = 7.42 \times 10^{-4} \text{ min.}^{-1}$ at 24.68°C , corresponding to $k_1 = 7.78 \times 10^{-4} \text{ min.}^{-1}$ at 25°C . The agreement is perfect.

Leininger & Kilpatrick (1938) studied the inversion of sucrose at 20 g./l. by hydrochloric acid over a wide range of concentrations and of temperatures, with the emphasis on determination of the temperature coefficient. At each temperature they found linearity between $\log(k_1/c)$ and c over a range of concentrations extending from 0.5 to 4 M. They express their results at each temperature by a formula of this type thus smoothing with respect to concentration, but they made no effort to smooth with respect to temperature. We find that their results over the temperature range 10°C to 40°C and over the concentration range 0.5 M to 4 M can be represented by the formula

$$\begin{aligned} \log_{10}(k_1/c) = & \bar{3}.880_8 + 0.197_8 c + (0.0640_7 - 0.0012_6 c)(t - 25) \\ & - (0.00025_5 - 0.00002_2 c)(t - 25)^2, \end{aligned}$$

where t denotes temperature on the Celsius scale. The agreement of their experimental results with this formula is better than 1 % except for four experiments out of about thirty. One of these discrepancies is at the high molarity 4.6; in another the discrepancy only just exceeds 1 %. The other two both at 35°C may possibly be interpreted as a temperature error of less than 0.1°C .

TABLE 5. FRÖKEN MOLIN'S MEASUREMENTS AT 28°C

S denotes concentration of sucrose in g./l. c_H denotes molar concentration of HCl. c_K denotes molar concentration of KCl. c denotes molar concentration of total chloride. k_1 denotes first order rate constant and is calculated from $k_1/\text{min.}^{-1} = 1.06 \times 10^{-2} c_H(1 + 10^{-3} S) \times 10^{20}$.

S	c_H	c_K	c	$k_1 \times 10^4$ (exp.)	$k_1 \times 10^4$ (calc.)
25	0.125	0.000	0.125	14.7	14.4
25	0.125	0.125	0.250	15.4	15.2
25	0.125	0.250	0.375	15.9	16.1
50	0.031	0.000	0.031	3.48	3.50
50	0.031	0.125	0.156	3.73	3.71
50	0.031	0.250	0.281	3.92	3.93
50	0.031	0.500	0.531	4.31*	4.41
100	0.125	0.000	0.125	15.7	15.4
100	0.125	0.125	0.250	16.6	16.4
100	0.125	0.250	0.375	17.4	17.3
100	0.125	0.500	0.625	19.1	19.4
200	0.031	0.000	0.031	4.03	4.00
200	0.031	0.125	0.156	4.33	4.24
200	0.031	0.250	0.281	4.61	4.49
200	0.031	0.500	0.531	5.11	5.04

Using the above formula to extrapolate down to a molarity 0.1 we obtain at 25°C

$$\log_{10} k_{\text{min.}} = \bar{4}.9006 \quad \text{or} \quad k_1 = 7.95 \times 10^{-4} \text{ min.}^{-1},$$

while our experiments by a slight extrapolation down to 20 g./l. of sucrose gave $7.32 \times 10^{-4} \text{ min.}^{-1}$ at 24.68°C corresponding to $7.67 \times 10^{-4} \text{ min.}^{-1}$ at 25°C . The

discrepancy of about $3\frac{1}{2}\%$ corresponds to a temperature difference of 0.2°C . Using the same formula again to extrapolate down to 0.1 M hydrochloric acid, we find

$$\frac{k \text{ at } 30^\circ\text{C}}{k \text{ at } 25^\circ\text{C}} = 2.058,$$

as compared with our experimental value 2.08 for which, however, we do not claim high accuracy.

One point that emerges from this review of the available experimental data is this. Since a change of 0.1°C in temperature corresponds to a $1\frac{1}{2}\%$ change in k , it follows that if an accuracy of 1% is claimed in the absolute value of k it is important to know the temperature with an accuracy and reliability of at least 0.05°C . We certainly cannot guarantee such a degree of certainty in the temperature of our own measurements and we hesitate to express any opinion on those of others, especially when nothing is stated about the thermometers used.

We are grateful to Professor F. G. Donnan, F.R.S., and to Professor C. K. Ingold, F.R.S., for the facilities provided for doing this work at University College, London.

REFERENCES

- Arrhenius, S. 1889. *Z. Phys. Chem.* **4**, 226.
 Bond, W. N. 1935 *Probability and random errors*, p. 112. London and New York: Longmans Green and Co.
 Bowe, L. E. 1927 *J. Phys. Chem.* **31**, 291.
 Brønsted, J. N. 1922a *Z. Phys. Chem.* **102**, 169.
 Brønsted, J. N. 1922b *J. Amer. Chem. Soc.* **44**, 877.
 Brønsted, J. N. 1926 *København's Universitets Festskrift*, p. 51.
 Debye, P. & Huckel, E. 1923 *Phys. Z.* **24**, 185.
 Euler, H. von 1900 *Z. Phys. Chem.* **32**, 348.
 Fales, H. A. & Morell, J. C. 1922 *J. Amer. Chem. Soc.* **44**, 2071.
 Floyd, W. W. 1931 *J. Phys. Chem.* **35**, 2968.
 Guggenheim, E. A. 1926 *Phil. Mag.* **2**, 538.
 Guggenheim, E. A. 1935 *Phil. Mag.* **19**, 588.
 Guggenheim, E. A. 1936 *Phil. Mag.* **22**, 322.
 Guggenheim, E. A. & Schindler, T. D. 1934 *J. Phys. Chem.* **38**, 543.
 Hitchcock, D. I. & Dougan, R. B. 1935 *J. Phys. Chem.* **39**, 1177.
 Kappanna, A. N. & Shrikhande, J. G. 1931 *J. Indian Chem. Soc.* **8**, 557.
 Kautz, C. P. & Robinson, A. L. 1928 *J. Amer. Chem. Soc.* **50**, 1022.
 La Mer, V. K. 1932 *Chem. Rev.* **10**, 179.
 Leminger, P. M. & Kilpatrick, M. 1938 *J. Amer. Chem. Soc.* **60**, 2891.
 Livingston, R. 1930 *J. Chem. Educ.* **7** (12), 2899.
 Palmaer, W. 1897 *Z. Phys. Chem.* **22**, 492.
 Pennycuik, S. W. 1926 *J. Amer. Chem. Soc.* **48**, 6.
 Scatchard, G. 1926 *J. Amer. Chem. Soc.* **48**, 2259.
 Schmid, G. & Olsen, R. 1926 *Z. Phys. Chem.* **124**, 97.

The propagation of a sound pulse in the presence of a semi-infinite, open-ended channel. II

BY W. CHESTER *Mathematics Department, University of Leeds*

(Communicated by S. Goldstein, F.R.S.—Received 7 January 1949)

Further investigation of a problem originally considered by the author (Chester 1950*a*) is carried out. One of the results of the above-mentioned paper is a description of the disturbance produced inside a two-dimensional, open-ended channel when a transient sound wave approaches the open end. At large distances from the open end this disturbance may be regarded as a plane wave travelling into the channel, and expressions were obtained for the potential and velocity distributions in a limited region at the head of the wave. The present paper discusses the asymptotic behaviour of the disturbance at large distances from the wave front. It is found that, for an incident Heaviside unit pulse, the wave inside the channel also tends to behave like a unit pulse, the correction term being $O(1/s)$ where s is the distance from the wave front.

The asymptotic behaviour for an arbitrary incident pulse is also considered and illustrated with typical examples.

The results are used to estimate the proportion of energy which returns along the channel when the incident pulse is of finite duration.

INTRODUCTION

In two previous papers (Chester 1950*a, b*, hereafter designated I and II), the author discussed the diffraction effects which occur when a plane wave of sound approaches the open end of a two-dimensional channel. In I, which was concerned with a transient wave or pulse, it was found that inside the channel and at large distances from the open end, the resulting disturbance could be regarded as a plane wave travelling into the channel. The form of the potential in a limited region at the head of the returning wave was obtained, but a complete description beyond this region was prevented by the complexity of the algebra. It is the purpose of the present paper to discuss the behaviour of this returning wave at large distances from the wave front, but at points sufficiently far distant from the open end for the assumption that the wave is plane to remain valid. This will be done with the help of the results obtained in II, which dealt with the analogous problem for a harmonic wave, by viewing the results of II as Fourier transforms and inverting by well-known methods.

RECAPITULATION

As in I, the incident pulse approaching the channel from a direction θ_0 is denoted by

$$\mathfrak{F}\{t + r \cos(\theta - \theta_0)\} H\{t + r \cos(\theta - \theta_0)\},$$

where

$$H(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

and is the Heaviside unit function. The origin of co-ordinates is taken at the upper lip of the channel and such that the channel wall is $\theta = \pm \pi$ (figure 1). For simplicity

the sonic velocity is taken to be unity. The resulting wave which is propagated into the channel, when it has become sensibly plane, is denoted by $\mathfrak{E}\{(t-x), \theta_0\} H\{t-x\}$, the positive direction of x being along the channel wall (figure 1).

$$\text{incident wave} \\ \mathfrak{F}\{t+r \cos (\theta-\theta_0)\} H\{t+r \cos (\theta-\theta_0)\}$$

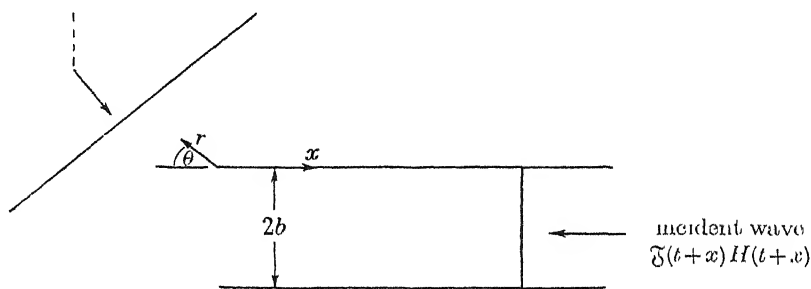


FIGURE 1

When the incident wave originates inside the channel the corresponding expressions are

$$\mathfrak{F}(t+x) H(t+x) \quad \text{and} \quad \mathfrak{B}(t-x) H(t-x).$$

The analysis in I is carried out in terms of the Heaviside operational calculus (Jeffreys & Jeffreys 1946; see also paper I, appendix), so that the above expressions become

$$F \exp [pr \cos (\theta-\theta_0)], \quad C \exp [-px], \quad F \exp [px], \quad B \exp [-px],$$

where F, C, B are the operational representations of $\mathfrak{F}(t), \mathfrak{E}(t), \mathfrak{B}(t)$ respectively, and the relation between $\mathfrak{F}(t)$ and $F(p)$, for example, is

$$\mathfrak{F}(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{e^{zt}}{z} F(z) dz. \quad (2)$$

It follows that if ϕ denotes the resulting potential field, then its asymptotic behaviour inside the channel is, for an incident external pulse,

$$\phi \sim C e^{-px}, \quad (3)$$

and for a pulse originating inside the channel,

$$\phi \sim F e^{px} + B e^{-px}. \quad (4)$$

APPLICATION OF THE HARMONIC SOLUTION

If the relation between $\mathfrak{F}(t)$ and its operational form F is modified by the substitution $z = -ik$, then equation (2) becomes

$$\mathfrak{F}(t) = -\frac{1}{2\pi i} \int_{-\infty+i\gamma}^{\infty+i\gamma} F(-ik) \frac{e^{-ikt}}{k} dk, \quad (5)$$

the incident wave is then represented by $F \exp [-ikr \cos (\theta-\theta_0)]$ (or $F \exp [-ikx]$) and the asymptotic behaviour of the final potential will be

$$\left. \begin{aligned} \phi &\sim C e^{ikx}, \\ \phi &\sim F e^{-ikx} + B e^{ikx}, \end{aligned} \right\} \quad (6)$$

or

corresponding to equations (3) and (4).

These expressions are now comparable with the relations which occur in the problem of the diffraction of a harmonic wave (cf. equations (2) and (3) of II with A replaced by F and the sign of x reversed). Although in the harmonic case k is a real parameter, the analysis in II was in fact carried out assuming that k has a positive imaginary part which was eventually taken to be zero. It follows that the relations for B and C , which are readily deduced from the results contained in II, should, when interpreted by (5), yield the answer to the transient problem. Unfortunately, the author has been unable to effect this inversion completely, but the results can be used to discuss the behaviour at a large distance behind the head of the returning wave. This is the purpose of the present investigation.

An expression for C for the case $F = 1$ is readily deduced from a combination of the relations (14), (24), (25), (62), (63), (65) and (70) of II. For an arbitrary initial wave an additional factor $F(-ik)$ is required, and the final result is found to be, for $-\pi < \theta_0 < \pi$,

$$C(\theta_0) = F(-ik) \prod_{n=1}^{\infty} \left\{ 1 - \frac{2b^2 k^2}{n^2 \pi^2} \sin^2 \frac{1}{2} \theta_0 - \frac{2ibk}{n\pi} \left(1 - \frac{k^2 b^2}{n^2 \pi^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2} \theta_0 \right\} \exp \left[\frac{2ibk}{n\pi} \sin^2 \frac{1}{2} \theta_0 \right] \\ \times \exp \left[\frac{2ibk \sin^2 \frac{1}{2} \theta_0}{\pi} \left(1 - \gamma + \log \frac{2\pi i}{bk} \right) - \frac{ibk \theta_0 \sin \theta_0}{\pi} \right], \quad (7)$$

where $2b$ is the width of the channel and γ is Euler's constant. The term $\left(1 - \frac{k^2 b^2}{n^2 \pi^2} \right)^{\frac{1}{2}}$ is real and positive when k is real and $|k| < n\pi/b$. The logarithm is evaluated on the branch for which $\log 1 = 0$.

This result is valid for $-\pi < \theta_0 < \pi$. At $\theta_0 = \pi$ C is discontinuous, the expression for $C(\pi)$ is, from equations (14) and (16) of II,

$$C(\pi) = \frac{1}{2} \lim_{\theta \rightarrow \pi} C(\theta). \quad (8)$$

Finally, from equations (29), (30) of I,

$$B = -2C(\pi) = -\lim_{\theta \rightarrow \pi} C(\theta). \quad (9)$$

The corresponding relations for the transient pulse are deduced from the above expressions with the aid of equation (5). Thus

$$\mathfrak{C}\{(t-x), \theta\} = -\frac{1}{2\pi i} \int_{-\infty+i\gamma}^{\infty+i\gamma} C(\theta) \frac{e^{-ik(t-x)}}{k} dk. \quad (10)$$

The expressions for $\mathfrak{C}\{(t-x), \pi\}$ and $\mathfrak{B}(t-x)$ follow at once from equations (8) and (9).

The problem of the asymptotic behaviour of \mathfrak{B} and \mathfrak{C} at large distances from the wave front now reduces to an investigation of the asymptotic behaviour of the integral in (10) for large values of $(t-x)$.

ASYMPTOTIC SOLUTION

It is convenient first to consider the case $F = 1$ (the original transient pulse is then $H\{t + r \cos(\theta - \theta_0)\}$).

With this simplification, it follows from (7) that C has branch points at $k = \pm n\pi/b$ equally spaced along the real axis. To determine the asymptotic behaviour of the

integral in (10) the contour of integration is accordingly deformed into a circuit round each branch point, adjacent circuits being linked by a straight portion which proceeds parallel to and at a small distance δ from the real axis on the negative side (figure 2).

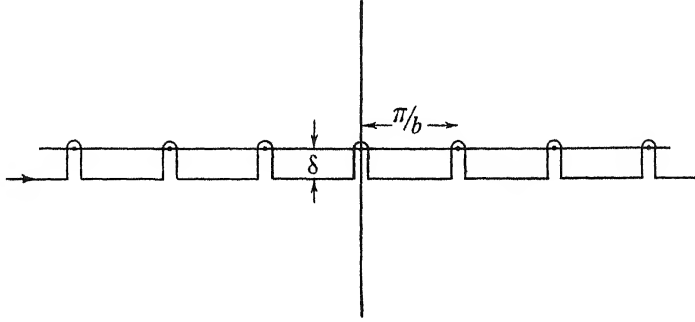


FIGURE 2. Contour of integration in the k -plane.

The integral along each of the straight portions parallel to the real axis is then exponentially small when $(t-x)$ is large.

The integral round the singularity at the origin is

$$1 - \frac{1}{\pi} \int_0^{2b\delta} \frac{\exp[-\eta(t-x)/2b]}{z} \sin(\eta \sin^2 \frac{1}{2}\theta) \prod_{n=1}^{\infty} \left\{ 1 + \frac{\eta^2}{2n^2\pi^2} \sin^2 \frac{1}{2}\theta - \frac{\eta}{n\pi} \left(1 + \frac{\eta^2}{4n^2\pi^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2}\theta \right\} \\ \times \exp \left[\frac{\eta \sin^2 \frac{1}{2}\theta}{n\pi} \right] \exp \left[\frac{\eta \sin^2 \frac{1}{2}\theta}{\pi} \left\{ 1 - \gamma + \log \frac{4\pi}{\eta} \right\} - \frac{\eta \theta \sin \theta}{2\pi} \right] d\eta \quad (11)$$

$$= 1 - \frac{1}{\pi} \int_0^{2b\delta} \exp[-\eta(t-x)/2b] \left[\sin^2 \frac{1}{2}\theta - \frac{\sin^4 \frac{1}{2}\theta}{\pi} \eta \log \eta + O(\eta) \right] d\eta \quad (12)$$

$$\sim 1 - \frac{\sin^2 \frac{1}{2}\theta}{\pi} \left(\frac{2b}{t-x} \right) - \frac{\sin^4 \frac{1}{2}\theta}{\pi^2} \left(\frac{2b}{t-x} \right) \log \left(\frac{t-x}{2b} \right). \quad (13)$$

The integral round the branch point at $k = m\pi/b$ ($m > 0$) is asymptotically

$$\left(\frac{1}{m} \right)^{\frac{1}{2}} \frac{i \sin^2 \frac{1}{2}\theta}{\pi^{\frac{1}{2}}} \left\{ \prod_{n=1}^{m-1} + \prod_{n=m+1}^{\infty} \right\} \left\{ 1 - \frac{2m^2}{n^2} \sin^2 \frac{1}{2}\theta - \frac{2im}{n} \left(1 - \frac{m^2}{n^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2}\theta \right\} \exp \left[\frac{2im}{n} \sin^2 \frac{1}{2}\theta \right] \\ \times \exp \left[2i \sin^2 \frac{1}{2}\theta + \frac{1}{4}i\pi + 2im \sin^2 \frac{1}{2}\theta \left(1 - \gamma + \log \frac{2i}{m} \right) - im\theta \sin \theta - 2im\pi \left(\frac{t-x}{2b} \right) \right] \\ \times \int_0^{2b\delta} \exp[-\eta(t-x)/2b] \eta^{\frac{1}{2}} d\eta \quad (14)$$

$$\sim A \left(\frac{2b}{t-x} \right)^{\frac{1}{2}} \exp \left[-2im\pi \left(\frac{t-x}{2b} \right) \right], \quad (15)$$

where A is independent of $(t-x)/2b$.

The behaviour is similar at the branch points on the negative real axis. It follows therefore, from relations (13) and (15), that in the particular case of a Heaviside unit pulse ($F = 1$),

$$\mathfrak{C}\{(t-x), \theta\} \sim 1 - \frac{\sin^2 \frac{1}{2}\theta}{\pi} \left(\frac{2b}{t-x} \right) + O \left(\frac{2b}{t-x} \right)^{\frac{3}{2}}, \quad (16)$$

if the integral occurring on the right-hand side of equation (10) is uniformly convergent along the deformed contour. This is shown to be true in the appendix.

The corresponding relation when the incident pulse originates inside the channel is, from equations (9) and (16),

$$\mathfrak{B}(t-x) \sim -1 + \frac{1}{\pi} \left(\frac{2b}{t-x} \right) + O \left(\frac{2b}{t-x} \right)^{\frac{3}{2}}. \quad (17)$$

ARBITRARY INITIAL PULSE

For an arbitrary initial pulse the investigation is restricted to the case when F remains finite as $k \rightarrow \infty$ on the contour of integration. The uniform convergence of the integral occurring in (10) then follows from the appendix with little modification. In addition, it is assumed that F has no singularities above the real axis. These conditions are not necessary, but seem to be satisfied for the majority of pulses having a physical significance. If, for example, $\mathfrak{F}(t)$ is finite and continuous for all $t > 0$, then from a well-known relation between \mathfrak{F} and F (Jeffreys & Jeffreys 1946), namely,

$$F(-ik) = -ik \int_0^\infty e^{ikt} \mathfrak{F}(t) dt, \quad (18)$$

it follows that F is regular for all k such that $\text{Im}[k] > 0$ (Whittaker & Watson 1946). This result will also remain valid if $\mathfrak{F}(t)$ contains a finite discontinuity.

If, in addition, F has no singularities on the real axis, then the contour of integration may be deformed as in figure 2. (This will be satisfied if $\mathfrak{F}(t) \sim e^{-\epsilon t}$ ($\epsilon > 0$) when t is large, since the right-hand side of equation (18) will then be regular for $\text{Im}[k] > -\epsilon$.) It follows, by an argument similar to that used in deriving equation (13), that the asymptotic behaviour of $\mathfrak{C}\{(t-x), \theta\}$ is in this case given by

$$\mathfrak{C}\{(t-x), \theta\} \sim F(0) \left\{ 1 - \frac{2b \sin^2 \frac{1}{2}\theta}{\pi(t-x)} \right\}, \quad (19)$$

unless $F(0) = 0$. In this latter case the asymptotic behaviour will, in general, be

$$\mathfrak{C} \sim -\frac{\sin^2 \frac{1}{2}\theta}{\pi} \int_0^\infty F(-\eta/2b) \exp[-\eta(t-x)/2b] d\eta + O \left(\frac{2b}{t-x} \right)^{\frac{3}{2}}. \quad (20)$$

If, for example, $F(-\eta/2b) \sim \eta^\alpha$ as $\eta \rightarrow 0$, then

$$\left. \begin{aligned} \mathfrak{C} &\sim -(\alpha!) \frac{\sin^2 \frac{1}{2}\theta}{\pi} \left(\frac{2b}{t-x} \right)^{\alpha+1} & (0 < \alpha < \tfrac{1}{2}) \\ &\sim O \left(\frac{2b}{t-x} \right)^{\frac{3}{2}} & (\alpha > \tfrac{1}{2}). \end{aligned} \right\} \quad (21)$$

It may also happen, for certain pulses, that $F = 0$ at $k = m\pi/b$ for all integral m . In this case equation (20) will require further modification if $\alpha > \frac{1}{2}$, since the asymptotic behaviour of the integral round the branch points at $k = m\pi/b$ will no longer be given by a relation similar to (15). A typical example is given by the finite pulse for which

$$\left. \begin{aligned} \mathfrak{F}(t) &= H(t) - H(t-2b), \\ F(-ik) &= 1 - e^{2ikb}. \end{aligned} \right\} \quad (22)$$

The asymptotic behaviour of C is now given by the first of equations (20), with $\alpha = 1$ and a change of sign, so that

$$\mathfrak{C} \sim \frac{\sin^2 \frac{1}{2}\theta}{\pi} \left(\frac{2b}{t-x} \right)^2 + O\left(\frac{2b}{t-x} \right)^{\frac{3}{2}}, \quad (23)$$

and in general it would seem that a finite pulse whose length is an integral multiple of the width of the channel produces a disturbance inside the channel which, at large distances from its front, tends to zero faster than the disturbance produced by a finite pulse of any other width.

Finally, when F has singularities on the real axis, the asymptotic expansion of \mathfrak{C} will contain extra terms from the deformation of the contour necessary to ensure that such singularities remain beneath the contour of integration. If, for example,

$$\left. \begin{aligned} \mathfrak{F}(t) &= e^{-i\hbar t}, \\ F &= \frac{k}{k-\hbar}, \end{aligned} \right\} \quad (24)$$

then

$$\begin{aligned} \mathfrak{C}\{(t-x), \theta\} &\sim \prod_{n=1}^{\infty} \left\{ 1 - \frac{2b^2\hbar^2}{n^2\pi^2} \sin^2 \frac{1}{2}\theta - \frac{2ib\hbar}{n\pi} \left(1 - \frac{b^2\hbar^2}{n^2\pi^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2}\theta \right\} \exp \left[\frac{2ib\hbar}{n\pi} \sin^2 \frac{1}{2}\theta \right] \\ &\times \exp \left[\frac{2ib\hbar}{\pi} \sin^2 \frac{1}{2}\theta \left(1 - \gamma + \log \frac{2\pi i}{b\hbar} \right) - \frac{ib\hbar \theta \sin \theta}{\pi} - i\hbar(t-x) \right] + O\left(\frac{2b}{t-x} \right)^{\frac{3}{2}}, \end{aligned} \quad (25)$$

$$\text{where, for } \left| \frac{b\hbar}{n\pi} \right| > 1, \quad \left(1 - \frac{b^2\hbar^2}{n^2\pi^2} \right)^{\frac{1}{2}} = -i \operatorname{sgn}[\hbar] \left(\frac{\hbar^2 b^2}{n^2\pi^2} - 1 \right)^{\frac{1}{2}} \quad (26)$$

$$\text{and} \quad \log \left(\frac{2\pi i}{b\hbar} \right) = \log \left(\frac{2\pi}{b|\hbar|} \right) + \frac{i\pi}{2} \operatorname{sgn}[\hbar]. \quad (27)$$

The first term in (25) is directly comparable with the expression obtained in II for the disturbance inside the channel produced by an incident harmonic wave. It should be noted, however, that whereas in II the result was restricted to values of \hbar within the range $0 < b\hbar < \frac{1}{2}\pi$, the above result is true for any value of \hbar . This is due to the different methods of proceeding to the limit in the transient and harmonic cases. In the former the returning wave is discussed at points which are at an infinite distance from the mouth of the channel and a finite distance from the head of the wave ($t, x \rightarrow \infty$, $(t-x)$ finite), while the results for the harmonic wave hold at a large but *fixed* distance from the mouth of the channel after an infinite time. It is shown in II that non-damped overtones may be propagated in the latter case if $b\hbar > \frac{1}{2}\pi$.

If the singularities in F are of higher order than one, then it will be necessary to separate the principal part of such singularities and integrate this separately before deforming the contour, but it is doubtful whether this case will be of practical significance.

ENERGY CONSIDERATIONS

In paper I the energy in the returning wave was discussed when the incident pulse originated inside the channel and the velocity distribution was given by a simple step function of length equal to the width of the channel, so that

$$\mathfrak{F}(t+x) = H(t+x) - H(t+x-2b). \quad (28)$$

The proportion of energy in the returning wave was calculated as a function of the distance from the wave front as far as the theory would allow. It was found that the part of the wave lying within a region of length $8b$ at the head of the wave (the limit of the calculations) contained 0.2788 of the total energy in the incident pulse and the energy contained in a length s at the head of the wave seemed to be approaching a limiting value in the region of 0.28.

The results now obtained enable an estimate to be made of the energy contained in that part of the wave beyond $(t-x) = 8b$. The calculations in I show that the velocity falls to a value -0.0633 at $(t-x) = 8b$, while from equations (9) and (23), its asymptotic behaviour is given by

$$-\frac{1}{\pi} \left(\frac{2b}{t-x} \right)^2, \quad (29)$$

when $(t-x)/2b$ is large. Since (29) takes the value 0.0199 for $(t-x) = 8b$, it would appear that further terms are required in the asymptotic expansion for an adequate representation of the velocity at this point. However, even if the velocity is assumed to tend to zero as slowly as

$$-K \left(\frac{2b}{t-x} \right)^2 \quad (30)$$

beyond $(t-x) = 8b$, where K ($= 1.0128$) is chosen to make the velocity continuous at this point, then the limiting value, as $(t-x) \rightarrow \infty$, for the proportion of the energy in the returning wave differs from its value for $(t-x) = 8b$ only by an amount 0.0053. It would seem, therefore, that the correct figure lies between 0.279 and 0.284.

If (30) is replaced by

$$-\frac{1}{\pi} \left(\frac{2b}{t-x} \right)^2 + K' \left(\frac{2b}{t-x} \right)^1, \quad (31)$$

where K' is determined by the continuity of the velocity, then the upper limit is reduced to 0.283.

APPENDIX

The validity of the asymptotic expansions derived in the report depends on the uniform convergence of the integral

$$\int \frac{C \exp[-ik(t-x)]}{ik} dk, \quad (32)$$

where the contour of integration is shown in figure 2, and C is given by equation (7) with $F = 1$.

It will first be shown that for large $|k|$ the modulus of the integrand at a point on the contour is of the same order as its value on the real axis at a point having equal real part. It is then shown that, on the real axis, the modulus of the integrand is $O(1/k^2)$ as $|k| \rightarrow \infty$. The uniform convergence of the integral in (32) then follows immediately.

The infinite product contained in C , namely,

$$\prod_{n=1}^{\infty} \left\{ 1 - \frac{2z^2}{n^2} \sin^2 \frac{1}{2}\theta - \frac{2iz}{n} \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2}\theta \right\} \exp \left[\frac{2iz}{n} \sin^2 \frac{1}{2}\theta \right] \quad (33)$$

(see equation (7) with $z = bk/n\pi$), may be written

$$\prod_{n=1}^{\infty} \left\{ -\frac{iz}{n} + \left(1 - \frac{z^2}{n^2}\right)^{\frac{1}{2}} \right\} \left(\frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2}\right)^{\frac{1}{2}} \right) \exp \left[\frac{iz}{n} (1 - \cos \theta) \right], \quad (34)$$

and since

$$\begin{aligned} \prod_{n=1}^{\infty} \left\{ \frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2}\right)^{\frac{1}{2}} \right\} \exp \left[-\frac{iz}{n} \cos \theta \right] \prod_{n=1}^{\infty} \left\{ -\frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2}\right)^{\frac{1}{2}} \right\} \exp \left[\frac{iz}{n} \cos \theta \right] \\ = \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2} \sin^2 \theta \right) = \frac{\sin(\pi z \sin \theta)}{\pi z \sin \theta}, \end{aligned} \quad (35)$$

it is sufficient to discuss the asymptotic behaviour of

$$\prod_{n=1}^{\infty} \left\{ -\frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2}\right)^{\frac{1}{2}} \right\} \exp \left[\frac{iz}{n} \cos \theta \right] \quad (36)$$

for $0 \leq \theta \leq \phi < \frac{1}{2}\pi$ and $z = a - i\delta$, where δ is small and $|a| \rightarrow \infty$. (The behaviour for $\theta = \frac{1}{2}\pi$ follows at once from equation (35).)

The investigation will be carried out explicitly when a is large and positive.

Consider first

$$\prod_{n=1}^{[a-1]} \left\{ \left(\frac{z^2}{n^2} - 1 \right)^{\frac{1}{2}} + \frac{z}{n} \cos \theta \right\}, \quad (37)$$

where $[a-1]$ denotes the largest integer smaller than $(a-1)$.

Under the above conditions, it can be shown that

$$\begin{aligned} \prod_{n=1}^{[a-1]} \left| \left(\frac{z^2}{n^2} - 1 \right)^{\frac{1}{2}} + \frac{z}{n} \cos \theta \right| &= \prod_{n=1}^{[a-1]} \left\{ \left(\frac{a^2}{n^2} - 1 \right)^{\frac{1}{2}} + \frac{a}{n} \cos \theta \right\} \left\{ 1 + O\left(\frac{a\delta^2}{(a^2 - n^2)^{\frac{1}{2}}} \right) \right\} \\ &= \{1 + O(\delta^2)\} \prod_{n=1}^{[a-1]} \frac{a}{n} \left\{ \left(1 - \frac{n^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\}. \end{aligned} \quad (38)$$

$$\text{Now} \quad \prod_{n=1}^{[a-1]} \left\{ \left(1 - \frac{n^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\} = \exp \sum_{n=1}^{[a-1]} \log \left\{ \left(1 - \frac{n^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\}, \quad (39)$$

and $\log \left\{ \left(1 - \frac{n^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\}$ is a monotonically decreasing function of n in the range considered so that

$$\sum_{n=1}^{[a-1]} \log \left\{ \left(1 - \frac{n^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\} = \int_0^a \log \left\{ \left(1 - \frac{x^2}{a^2} \right)^{\frac{1}{2}} + \cos \theta \right\} dx + O(1) \quad (40)$$

$$= a \log \cos \theta + a \left\{ \frac{1}{2}\pi \cos \theta - 1 \right\} + \frac{1}{2}a \sin \theta \log \left(\frac{1 + \sin \theta}{1 - \sin \theta} \right) + O(1). \quad (41)$$

It follows immediately from (41) and the use of Stirling's formula, namely,

$$(a!) \sim e^{-a} a^a (2\pi a)^{\frac{1}{2}}, \quad (42)$$

that

$$\prod_{n=1}^{[a-1]} \left| \left(\frac{z^2}{n^2} - 1 \right)^{\frac{1}{2}} + \frac{z}{n} \cos \theta \right| \sim \frac{K}{a^{\frac{1}{2}}} \exp \left[a \log \cos \theta + \frac{1}{2}\pi a \cos \theta + \frac{1}{2}a \sin \theta \log \left(\frac{1 + \sin \theta}{1 - \sin \theta} \right) \right], \quad (43)$$

where, here and henceforth, K denotes *some* constant factor (which may take different values when used in different relations).

In a similar manner it can be shown that

$$\prod_{n=[a+2]}^{\infty} \left| \left\{ \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} - \frac{iz}{n} \cos \theta \right\} \exp \left[\frac{iz}{n} \cos \theta \right] \right| = \prod_{n=[a+2]}^{\infty} \left\{ 1 - \frac{a^2}{n^2} \sin^2 \theta \right\}^{\frac{1}{2}} \left\{ 1 + O\left(\frac{a\delta}{n^2}\right) \right\} \\ = \{1 + O(\delta)\} \prod_{n=[a+2]}^{\infty} \left\{ 1 - \frac{a^2}{n^2} \sin^2 \theta \right\}^{\frac{1}{2}}. \quad (44)$$

$$\text{Now } \prod_{n=[a+2]}^{\infty} \left\{ 1 - \frac{a^2}{n^2} \sin^2 \theta \right\} = \frac{\{([a+2]-1)!\}^2}{([a+2]+a \sin \theta - 1)! ([a+2]-a \sin \theta - 1)!} \\ \sim K \exp \left[-2a \log \cos \theta - a \sin \theta \log \left(\frac{1+\sin \theta}{1-\sin \theta} \right) \right], \quad (45)$$

so that

$$\prod_{n=[a+2]}^{\infty} \left| \left\{ \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} - \frac{iz}{n} \cos \theta \right\} \exp \left[\frac{iz}{n} \cos \theta \right] \right| \\ \sim K \exp \left[-a \log \cos \theta - \frac{a \sin \theta}{2} \log \left(\frac{1+\sin \theta}{1-\sin \theta} \right) \right]. \quad (46)$$

With the aid of the results expressed in (43) and (46), together with the relation

$$\left| \prod_{n=1}^{[a-1]} \exp \left[\frac{iz}{n} \cos \theta \right] \right| = \prod_{n=1}^{[a-1]} \exp \left[\frac{\delta}{n} \cos \theta \right] \sim a^{\delta \cos \theta}, \quad (47)$$

it follows that, for $0 \leq \theta \leq \phi < \frac{1}{2}\pi$,

$$\left| \prod_{n=1}^{\infty} \left\{ -\frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} \right\} \exp \left[\frac{iz}{n} \cos \theta \right] \right| \sim \frac{K a^{\delta \cos \theta}}{a^{\frac{1}{2}}} \exp \left[\frac{1}{2} \pi a \cos \theta \right]. \quad (48)$$

In particular, for $\theta = 0$

$$\left| \prod_{n=1}^{\infty} \left\{ -\frac{iz}{n} + \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} \right\} \exp \left[\frac{iz}{n} \right] \right| \sim K a^{\delta - \frac{1}{2}} \exp \left[\frac{1}{2} \pi a \right] \quad (49)$$

Furthermore, a combination of relations (35) and (48) shows that, for $0 \leq \theta \leq \phi < \frac{1}{2}\pi$,

$$\left| \prod_{n=1}^{\infty} \left\{ \frac{iz}{n} \cos \theta + \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} \right\} \exp \left[-\frac{iz}{n} \cos \theta \right] \right| \sim \frac{K \sin(\pi a \sin \theta)}{\pi a^{\frac{1}{2}} \sin \theta} a^{-\delta \cos \theta} \exp \left[-\frac{1}{2} \pi a \cos \theta \right]. \quad (50)$$

Finally, when $\theta = \frac{1}{2}\pi$, the asymptotic behaviour is obtained immediately from equation (35). Explicitly it is

$$\left[\frac{\sin(\pi a \sin \theta)}{\pi a \sin \theta} \right]^{\frac{1}{2}}. \quad (51)$$

It follows from relations (48) to (51), together with the equality of the expressions (33) and (34), that for $0 \leq \theta \leq \pi$ and a large and positive,

$$\left| \prod_{n=1}^{\infty} \left\{ 1 - \frac{2z^2}{n^2} \sin^2 \frac{1}{2} \theta - \frac{2iz}{n} \left(1 - \frac{z^2}{n^2} \right)^{\frac{1}{2}} \sin^2 \frac{1}{2} \theta \right\} \exp \left[\frac{2iz}{n} \sin^2 \frac{1}{2} \theta \right] \right| \\ = O[a^{2\delta \sin^2 \frac{1}{2} \theta - 1} \exp \{ \pi a \sin^2 \frac{1}{2} \theta \}]. \quad (52)$$

It is now a simple matter to obtain the behaviour of $|C|$ at points on the contour for which $\text{Re}[k]$ is large and positive. If equation (52) is substituted in equation (7), the result is, for $F = 1$,

$$|C| = O\left(\frac{1}{|k|}\right), \quad (53)$$

so that the modulus of the integrand in expression (32) is $O\left(\frac{1}{|k|^2}\right)$ when $\text{Re}[k]$ is large and positive. A similar result can be obtained for $\text{Re}[k]$ large and negative, from which it follows that the integral converges uniformly for all $(t-x) > 0$.

REFERENCES

- Chester, W. 1950a *Phil. Trans. A* **242**, 527
 Chester, W. 1950b *Phil. Mag* **41**, 11.
 Jeffreys, H. & Jeffreys, B. S. 1946 *Functions of mathematical physics*. Cambridge University Press
 Whittaker, E. T. & Watson, G. N. 1946 *Modern analysis*, 4th ed. p. 93. Cambridge University Press.

Measurement of the adsorption of surface-active agents at a solution/air interface by a radiotracer method

BY D. J. SALLEY, A. J. WEITH, JR., ANN A. ARGYLE AND J. K. DIXON

*Stamford Research Laboratories, American Cyanamid Company,
 Stamford, Connecticut, U.S.A.*

(Communicated by E. K. Rideal, F.R.S.—Received 23 January 1950—
 Revised 3 April 1950)

The adsorption of a water-soluble surface-active agent at the solution/air interface has been measured by taking advantage of the soft beta-radiation from the radioactive isotope of sulphur, ^{35}S . The method depends upon the fact that, because of 'self-absorption' of the radiation by the solution, and because the molecules of a surface-active agent are preferentially adsorbed at the interface, the radioactivity above a solution containing such an agent labelled with radio-sulphur should be higher than that from a solution (e.g. of sodium sulphate) in which no surface adsorption occurs.

Carefully purified di-*n*-octyl sodium sulposuccinate and sodium sulphate were each synthesized with ^{35}S . Measurements on solutions of these showed that the radiotracer technique permitted a successful quantitative determination of the adsorption. Counting rates over the solutions of agent were 1.1 to 20 times those of the corresponding sodium sulphate, and for some 29 different solutions covering a range of specific activities (1×10^{12} to 5×10^9) counts/min./mol.) and of concentrations (0.1 to 150×10^{-3} mol./ml.) a well-defined adsorption isotherm resulted. The surface excess obtained from the tracer measurements agreed with that calculated from surface-tension measurements by the Gibbs equation if it was assumed that no Na^+ and only H^+ were associated with the long-chain ions adsorbed in the interface. The radioactivity measurements suggested further that, after the agent had formed a unimolecular layer, a further increase in bulk concentration caused further adsorption at the interface.

An interesting aspect is that the rate of adsorption was followed in the low concentration ranges. The method promises to be useful for studying the kinetics of such adsorption processes.

The radiotracer method outlined above seems to be generally applicable to surface phenomena involving adsorbed layers, either mono- or multimolecular in nature. Although this paper describes the use of ^{35}S , it is evident that other radioisotopes having sufficiently soft radiation, such as ^{14}C , ^{45}Ca , ^{59}Fe , etc., could be employed. Moreover, the method appears to open considerable possibility for examining interaction in surface layers.

INTRODUCTION

The problem of measuring the amount of water-soluble surface-active agent adsorbed at the air/solution interface is of fundamental importance. Especially is this true in connexion with surface-tension studies based on the Gibbs adsorption equation. The air-bubble method (Donnan & Barker 1911; McBain & Dubois 1929) and the 'microtome' method (McBain & Humphreys 1932; McBain & Swain 1936) provided a limited number of direct measurements of surface adsorption, but no simple method of measuring adsorption over a wide range of concentration has been available. It is the purpose of this report to describe a procedure based upon the use of radiotracers. A preliminary note on this work has already appeared (Dixon, Weith, Argyle & Salley 1949).

Consider two solutions, one of ordinary electrolyte and one of surface-active agent. Let these solutions be identical with regard to bulk molar concentration, volume, area of air/water interface, etc., and further, let each contain equal numbers of molecules tagged with a radioelement. The 'self-absorption' of the radiation coming from within the solution will then be the same in both solutions, but in the case of the surface-active agent, the radiation coming from those molecules adsorbed at the interface will not be 'self-absorbed'. The activity above the surface of such a solution consequently will be higher than that from the solution of ordinary electrolyte, and the difference between the two will give the amount of surface adsorption.

In principle, then, the surface adsorption may be measured, but whether the measurement is experimentally feasible depends upon the relative magnitudes of the surface and the bulk radiation intensities (equation (1)). Of prime importance in this regard is the penetrating power of the radiation, since the softer the radiation the more feasible is the measurement; such soft beta emitters as ^{35}S and ^{14}C are particularly suitable. However, the concentration of the solution is also concerned and the relationships of the various factors involved will be brought out in the following development.

For a solution of surface-active agent

$$I_{\text{observed}} = I_{\text{surface}} + I_{\text{bulk}}, \quad (1)$$

where the I 's represent intensities of radiation. The same equation holds for an ordinary electrolyte except that $I_{\text{surface}} = 0$.

Let I' be the activity in counts per minute coming from a solution of a non-surface-active electrolyte of given depth in a cylindrical shallow vessel of given area, as observed with a mica-window counter tube in position over the vessel. Let I'_0 be the activity which the electrolyte would show if no absorption of the radiation occurred in the solution. The ratio $(I'/I'_0) = f$ then represents the fraction of the radiation not absorbed by the solution. This is an empirical value dependent on the softness of the radiation and on the experimental arrangement. It is independent of concentration as long as the densities of the solutions are essentially constant. It is independent also of the molecular species tagged with the radioisotope.

Let I_0 be the intensity which the bulk solution of a surface-active agent would show if no self-absorption occurred. Then

$$I_0 = mvs, \quad (2)$$

where m , in mol./ml., is the equilibrium bulk concentration of the agent after adsorption equilibrium is reached, v is the volume in ml. of solution, and s is the specific activity of the agent in counts/min./mol. As a result of self-absorption, the intensity from the solution will be reduced by the term f so that

$$I_{\text{bulk}} = I_0 f = mvsf. \quad (3)$$

Let m_a be the mols adsorbed at the solution/air interface

$$I_{\text{surface}} = m_a s. \quad (4)$$

The initial concentration of the agent, m_0 , is decreased because of the adsorption of agent in the surface, so that at equilibrium

$$mv = m_0 v - m_a. \quad (5)$$

Substituting (3), (4) and (5) in (1)

$$I_{\text{observed}} = m_a s + (m_0 v - m_a) sf, \quad (6)$$

and solving for m_a

$$m_a = \frac{I_{\text{observed}} - m_0 v sf}{s(1 - f)}. \quad (7)$$

It is clear from equation (7) that the smaller the value of f , the more feasible is the experimental measurement. Further, bearing in mind the form of the usual adsorption isotherm, the precision of measurement becomes better the smaller the value of m_0 at which the 'monolayer' is complete. Finally, equation (7) shows that an accurate value for m_a depends on an accurate determination of the specific activity term, s . The latter, strictly speaking, represents the number of counts per minute a mol. of the substance would show when spread uniformly over the surface of the water in the geometrical arrangement at hand with no self-absorption occurring.

The surface-active agent, di-*n*-octyl sodium sulphosuccinate labelled with ^{35}S , was chosen as experimental material. Calculations predicted that the absorption coefficient of the beta-radiation from this isotope was well suited for the purpose, and that the ^{35}S available from the U.S. Atomic Energy Commission was of sufficiently high specific activity. Knowledge of the surface tension of solutions of the sulphosuccinate indicated that the monolayer should be complete at relatively low concentrations. Furthermore, synthesis with ^{35}S and subsequent purification seemed fairly straightforward.

EXPERIMENTAL

Preparation of di-n-octyl radiosulphosuccinate

The synthesis of di-*n*-octyl sodium sulphosuccinate on a microscale with radio-sulphur involved precipitation of ^{35}S as $\text{Ba}^{35}\text{SO}_4$, reduction to $^{35}\text{SO}_2$, reaction with Na_2CO_3 to form $\text{NaH}^{35}\text{SO}_3$, and reaction of the latter with di-*n*-octyl maleate.

Radiosulphur, ^{35}S , formed by neutron bombardment of potassium chloride,* was precipitated as $\text{Ba}^{35}\text{SO}_4$ from hot aqueous solution. The active precipitate (12 mg.)

* The material was obtained from the Oak Ridge National Laboratories on allocation from the U.S. Atomic Energy Commission.

was reduced with an equal weight of finely powdered iron at 1200° C for 1 hr in a stream of dry nitrogen, the $^{35}\text{SO}_2$ (98 % yield) being caught in a U-trap in liquid air.* Di-*n*-octyl maleate (twice fractionally distilled, b.p. 169° C at less than 1 mm.) was dissolved in absolute ethanol to give 0.002 mol. of ester/g. of solution. Sodium carbonate (3.3 mg. as a 10 % solution) was deposited by freeze-drying as a thin layer of solid on the walls of a small ampoule. Water (6.8 mg.) and the ester-alcohol solution (31.0 mg.) were added. After freezing and evacuating this system as well as the U-tube containing the $^{35}\text{SO}_2$, the gas was distilled under vacuum into the ampoule. The ampoule was then rotated about 8 hr. in an oil bath at 80° C to produce the radiosulphosuccinate.

For purification purposes, the contents of the reaction ampoule were vacuum-dried from the frozen state and extracted three times with carbon tetrachloride, rejecting insoluble inorganic material. The combined extracts were filtered, then carefully pumped off, leaving a radioactive solid (18.5 mg.), to which was added 52.5 mg. of inactive sulphosuccinate as carrier. The whole was dissolved in just sufficient 75/25 alcohol/water mixture to effect complete solution at room temperature. Recrystallization was then carried out in a small centrifuge at 1° C. Five recrystallizations were made, the final weight of product being approximately 10 mg. Between the second and fifth recrystallization the specific activity increased only 10 %, from 1.26 to 1.38×10^{12} counts/min./mol.; this is a fair criterion of purity of the final product. The yield calculated from the specific activities of the initial and final materials and the known amount of carrier was about 10 %.

Preparation of solutions

A master solution of radioactive sulphosuccinate was prepared from dry material, accurately weighed (microbalance) into a 1 ml. microflask. From this solution a stock solution was prepared as desired by pipetting an aliquot into a 1 ml. microflask, and in turn from this stock solution aliquots were withdrawn just before use and diluted to 10 ml. to provide the solutions used in the actual adsorption measurements. These steps were found conducive to reproducible results. The master solution was kept in dry ice to avoid possible decomposition, and no solution was at room temperature for more than 8 hr. Each solution was used only once for an adsorption measurement.

Activity measurements

For activity measurements, a conventional bell-shaped mica window counter tube (1.43 cm. inside radius) having a window of about 2 mg./cm.² thickness was used in conjunction with a standard scaling circuit. The tube was routinely monitored with a standard uranium, a standard ^{14}C , and a sodium sulphate sample.

For counting solutions, a small stainless steel vessel (0.58 cm. inside depth, 1.62 cm. inside radius) was fitted into a holder by which it could be centred reproducibly under the counter tube. Ordinarily 2 ml. of solution were placed in the vessel, making a solution depth of 0.242 cm. with the surface 0.69 cm. from the mica window. As expected, solution depth was not critical, the observed activity of sulphate solutions increasing only 8 % when the volume was increased from 2 to 3 ml.

* The procedure was developed by R. K. Madison with one of us (D.J.S.).

For counting dried aliquots of solution, two methods were employed. In one of these, aluminium foils (5 mg./cm.² thick), of the same diameter as the inside of the vessel, were spotted with tiny droplets (40 to 50 for a 0.1 ml. aliquot) as closely together as possible over the entire area. These, when dried under an infra-red lamp, yielded a sample covering fairly well the whole surface. To position the foil under the counter tube, it was floated on water inside the vessel or built up on aluminium foil to the usual level of the solution. This procedure was thought to reproduce with considerable exactness the geometry of the surface of the solution, and is of importance in the determination of specific activity.

The above method of counting dried aliquots was not convenient for routine work, so that in most cases an aliquot (0.1 ml.) of solution was dried as a single drop in a small metal cup. Multiplication of the count in the cup by 0.905 gave the count on the aluminium foil.

No self-absorption correction was necessary for the counts of the dried aliquots since care was exercised that the layer of dried solid was less than 0.1 mg./cm.² thick; this was possible because of the high specific activity of the preparations.

Determination of f

The quantity representing the absorption of ³⁵S radiation by water in the arrangement employed was obtained simply from the activity of 2 ml. of a given sulphate solution and of 0.1 ml. of a dried aliquot. The average value of 23 comparisons was $f = 0.0153 \pm 0.0004$. Using the usual exponential formula $f = \frac{1 - e^{-\alpha d}}{\alpha}$ and noting that d was 0.242 cm., an absorption coefficient, α , of 270 cm.⁻¹ for ³⁵S radiation in water was derived. This corresponds to a value for the half-thickness in water of 2.56 mg./cm.²; 2.65 mg./cm.² was measured for aluminium. Exact agreement is not to be expected since the absorption in water and aluminium would not be identical.

Specific activity

The specific activity was defined in the introduction as the activity in counts per minute which a mol. of the substance would show when spread uniformly over the surface of water under the geometrical conditions existing, with no self-absorption. An accurate determination of this quantity is necessary in order finally to obtain absolute values of the amount of agent adsorbed at an interface, but its experimental measurement was somewhat troublesome. Regarding the geometrical and self-absorption requirements, it was presumed, as mentioned above, that these were sufficiently well approximated by the technique of spotting the sample on an aluminium foil. However, the count from a sample spread on water is different from that for one on aluminium because of differences in backscattering. Correction for such backscattering was sought by two methods. In the one a sodium sulphate sample was dried on thin aluminium foil (3.3 mg./cm.²) and counted in an upside down position first in air, then over aluminium, and finally with a water surface placed as closely as possible under the sample but not touching it. The ratio of activities of air to aluminium was 0.91 ± 0.02 and that of water to air was 0.99 ± 0.02 . To attempt to eliminate the effect of the support, in the second method the active samples were

dried on thin films of flexible collodion 0.03 to 0.06 mg./cm.² thick, these were counted first over air and then when supported just above the water surface. The value for the ratio of water/air was 1.01 ± 0.02 , in agreement with the first method. The fact that the backscattering by water was essentially the same as that by air is in agreement with expectations from the observations of others regarding backscattering by various materials (Calvin, Heidelberger, Reid, Tolbert & Yankwich 1949). To obtain the specific activity of any sample counts/min./mol., therefore, the count observed for a dried aliquot spotted on an aluminium foil and counted at the surface of the water in the vessel was multiplied by 0.91 , and then divided by the known amount of material taken.

Ordinarily, aliquots of solutions of the sulphosuccinate were dried and counted to obtain specific activity measurements. To check upon the results so obtained, micro-Carius oxidation of the sulphosuccinate was carried out (Niederl & Niederl 1942) and the specific activity of the resulting sodium sulphate was determined. The values by the two methods agreed, but the precision was higher with the latter ($\pm 3\%$), so that the specific activity figure used throughout has been based on the micro-Carius procedure.

RESULTS

Adsorption isotherm by radioactivity method

The adsorption of the di-*n*-octyl sodium sulphosuccinate in the air/solution interface has been measured for twenty-nine solutions in seven series of runs, covering concentrations from 0.1 to 150×10^{-8} mol./ml. and specific activities from about 10^{12} to 10^{10} counts/min./mol. The temperature of all the measurements was that of the air-conditioned room 23.9°C .

The observed activities of the sulphosuccinate solutions were always higher than those expected for corresponding sulphate solutions. Table 1 records data for nine typical runs covering low and high concentration ranges. Column 2 gives the initial concentrations. Columns 3 and 4 show the specific activity as determined from Carius oxidation and from the sulphosuccinate solutions, respectively; these data will be discussed below. Columns 5 and 6 contain the calculated and the observed counts, respectively. The former is the second term of the numerator of equation (7); it represents the count which would have been shown by a sodium sulphate solution of the same concentration and specific activity as the sulphosuccinate solution. (It may be pointed out here that no correction was made for possible negative adsorption of the sulphate since all the concentrations were so low as to make any such effect entirely negligible.) Column 7 indicates the count in the entire surface layer of 8.25 cm.² area.

Column 8 contains the number of mols adsorbed per cm.²; these are calculated from equation (7). Column 9 gives the solution concentration at equilibrium, as computed by equation (5).

An isotherm showing the mols adsorbed per cm.² against equilibrium concentration is plotted in figure 1, from the data of table 1 along with similar data for all the other twenty runs. An enlargement of the initial portion of the isotherm is shown in figure 2.

The isotherm indicates that as the concentration increases the amount adsorbed in the surface layer rises rather steeply to about 2.3×10^{-10} mol./cm.² at an equilibrium concentration of about 4×10^{-8} mol./ml. From this point there is a gradual steady increase in the adsorption until approximately 5.5×10^{-10} mol. are adsorbed at the highest concentration measured.

TABLE 1. EXPERIMENTAL DATA ON RADIOACTIVE SODIUM DI-N-OCTYL SULPHOSUCCINATE SOLUTIONS

no.	initial conc , m_0 (mol /ml. $\times 10^8$)	specific activity counts/min./mol $\times 10^{-10}$		count (counts/min.)		count in surface (counts/ min.)	mols adsorbed per cm. ² (m_a /cm. ² $\times 10^{10}$)	equilibrium conc , m (mol./ml $\times 10^8$)
		from Carius oxidation	from direct measure- ment	calculated	observed			
38	1.3	150	100	518*	2800	2282	1.89	1.05
48.3	1.42	128	123	592	2959	2367	2.12	1.33
70.1	0.89	118	118	323*	2360	2037	2.13	0.808
78.1	0.09	118	108	33	710	677	0.709	0.061
78.3	0.90	118	120	325	2110	1785	1.87	0.823
82.4	30.7	8.77	8.66	824	1053	229	3.12	30.6
187.2	68.9	2.18	2.27	453	524	71	3.90	68.7
187.5	125.3	2.12	2.10	814	896	82	4.63	125.2
183.2	71.1	1.27	1.30	272	309	37	3.49	70.9

* These counts for the sulphate solutions were observed directly.

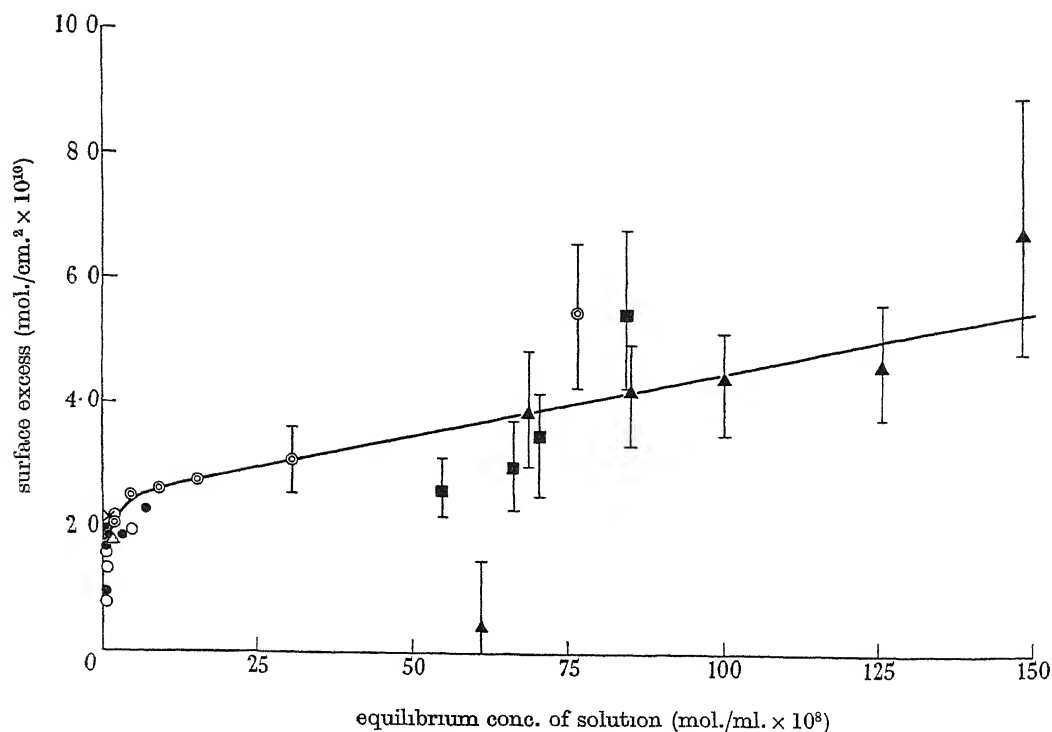


FIGURE 1. Adsorption isotherm as determined from radioactivity. Δ solution 38, \bullet series 48, \times solution 70.1, \circ series 78, \odot series 82, \blacksquare series 183, \blacktriangle series 187.

Enough counts on each sample were generally made to make the probable error in counting less than $\pm 2\%$. At low concentrations there is good agreement between runs, but the scattering is rather large at the higher concentrations because the difference between the observed and calculated counts then becomes relatively small. Actually, with the probable error of each count of only $\pm 1\%$, the difference in counts may be in error by as much as ± 15 to 25% . The vertical lines through the points at the higher concentrations have been estimated to illustrate this spread, and they indicate clearly that too much significance cannot be attached to the scattering.

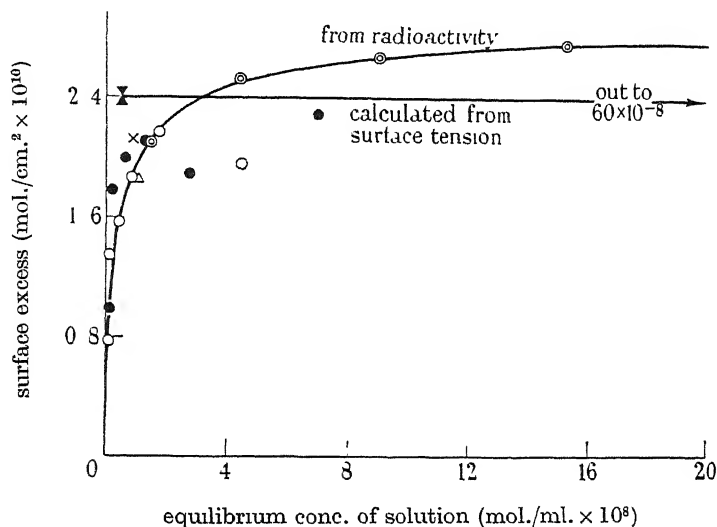


FIGURE 2. Enlargement of radioactivity isotherm. \triangle solution 38, \bullet series 48, \times solution 70-1, \circ series 78, \odot series 82.

The accuracy of the results is difficult to assess. It is limited primarily by the uncertainty in the value of the specific activity, which may be in error because of errors in the geometry and backscattering corrections. It is believed that these do not exceed 20% ; consequently the position of the isotherm on the graph cannot be much higher or lower than that shown.

In the following paragraphs, several points of general importance will be mentioned.

Solutions 38 and 70-1 represent two cases for which there was obtainable a direct comparison of the count for identical solutions of the sulphosuccinate and of sulphate produced by micro-Carius oxidation. They thus carry out the experiment proposed in the introduction.

To obtain the lower specific activity necessary in the runs at higher concentrations, active sulphosuccinate was added to inactive material. This latter had been prepared and recrystallized on a large scale, analysis for water, sulphur, sodium and ester content showing a purity of at least 99% . The very good agreement of the results after dilution of the active materials with the inactive is indicated by comparison of series 78 and 82 (figure 2), in which the highest concentrations of the former overlapped the lowest of the latter; the specific activity was different by about 15-fold.

Thus if any impurity was present in the active sulposuccinate, or in the water used in preparing the solution, the same kind and amount of impurity was present in both the active and inactive preparations.

The so-called Carius specific activity of column 3 of table 1 is based on the activity of the sulphate as determined after Carius oxidation of an aliquot of a weighed sample of the sulposuccinate. The figures of column 4 are based on counts of dried aliquots of the sulposuccinate solutions used in the actual surface adsorption measurements. If the concentrations of these solutions (as calculated from the concentration of the stock solution and the known dilutions) were correct, the specific activity of columns 3 and 4 should agree. The data of the table shows that this is true within a limit of error of $\pm 5\%$ except for solutions 38 and 48-3. The fact that the concentrations were correct shows that the aliquoting and other manipulations were satisfactory. It further indicates that adsorption of the sulposuccinate by the glass of the volumetric flask was not appreciable within the limit of error. In order to attain the agreement noted, however, it was found important that the 0.1 ml. aliquots be taken when the volumetric flask was full to the mark since this provided a low ratio for the area of the air/solution interface to volume. The disagreement shown by solutions 38 and 48-3 resulted from failure to adhere to this sequence.

Measurement of surface tension

Measurements of the surface tension of di-*n*-octyl sodium sulposuccinate solutions were made by the du Noüy ring method in its refined form as described by Harkins & Jordan (1930). The solutions were prepared from the purified un-

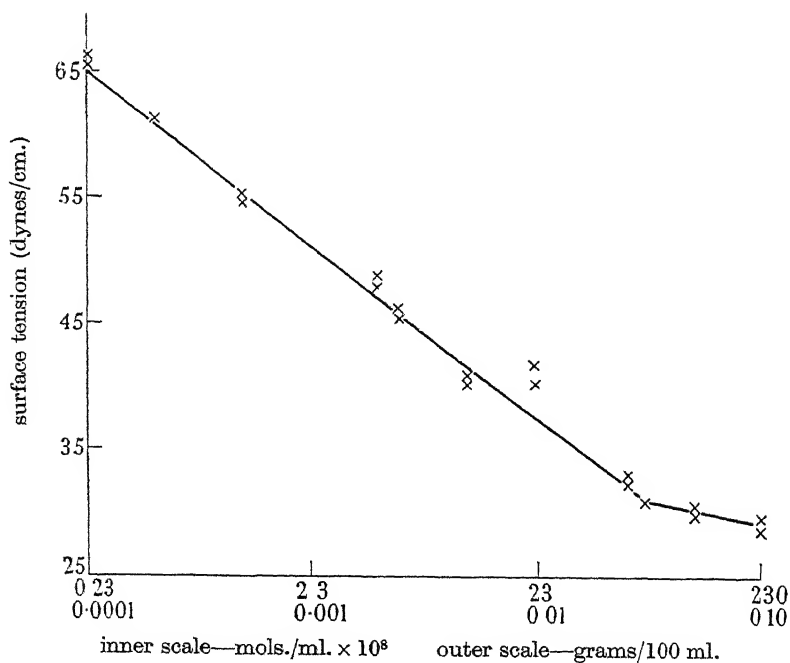


FIGURE 3. Surface-tension against concentration curve.

tagged agent mentioned previously. The concentrations ranged from 0.2×10^{-8} to 225×10^{-8} mol./ml., the lower limit being that for satisfactory precision. Surface tension measurements made on the solutions employed in the counting experiments were in substantial agreement with those for untagged agent.

The results of the surface tension measurements are shown in figure 3. The curve shows no minimum at the concentration where micelles are formed and gives satisfactorily the average slope of the surface tension-concentration curve over the range studied. However, there was indication of a slight concavity to the abscissa axis, but since the *slope* could not be established to within $\pm 20\%$ over any two- or threefold variation of concentration the best representation is the straight line shown.

Comparison of adsorption isotherms

The Gibbs adsorption isotherm expresses the relationship between the surface tension and the excess of material in the surface layer (Adam 1938). In the development which follows concentration, c , in mol./ml., is employed in place of solute activity as a first approximation; this assumption is justified here because the solutions are very dilute and because the activity coefficients are close to unity for certain homologues of the sulphosuccinate used (McBain & Bolduan 1943).

The development of the Gibbs equation for uni-univalent electrolytes completely dissociated in solution leads, according to Brady (1949), to the final form

$$\Gamma = -\frac{1}{2RT} \frac{d\gamma}{d \ln c}, \quad (8)$$

where γ is the surface tension, Γ is the excess adsorbed in the surface, and c is the bulk equilibrium concentration. This equation contains a factor of two which does not appear in the conventional Gibbs equation for a non-dissociating solute. The factor of two arises because of the assumption that the salt, in this case sodium sulphosuccinate, is adsorbed in the interface, the two ions being present in equal amounts.

In the present instance equation (8) does not appear to fit the experimental facts. The amount of sulphosuccinate adsorbed in the interface as calculated from the surface tension curve of figure 3 using equation (8) gives a value of about 1.2×10^{-8} mol./cm.² for the concentration range 0.2 to 70×10^{-8} mol./ml. This is about one-half the value observed by the radiotracer method, estimated from figure 2 as 2.4×10^{-8} mol./cm.² at the knee of the isotherm, and it is believed that the precision of both the counting and the surface-tension measurements is such that the difference is real. Another disturbing point is that the area for the sulphosuccinate molecule calculated from surface tension in this manner is about 130 \AA^2 , a value about twice as high as might be expected.

These serious discrepancies led to a re-examination of the assumptions upon which the above equation was developed. In a private discussion Professor L. Onsager of Yale University suggested that the capillary-active species adsorbed in the interface was the free *bis*-(*n* octyl-) sulphosuccinic acid, and not the sodium salt. This could arise from hydrogen bonding of the sulphosuccinate ions in the interface, and/or from the fact that the adsorption of the long-chain ions in the surface produces

a potential which would greatly lower the dissociation of the sulphosuccinic acid. In any event, the protons would be supplied by dissociation of the water in the bulk solution, the water acting as an infinite buffer. On the assumption that the sulphosuccinic acid is the adsorbed species and that the hydrogen-ion concentration does not change in the bulk of the solution, the Gibbs equation reduces to the conventional form

$$\Gamma = -\frac{1}{RT} \frac{d\gamma}{d \ln c}. \quad (9)$$

The factor of two in the equation is thus eliminated by the assumption that the acid is at the interface. The excess measured by the radiotracer method and that calculated by equation (9) becomes practically identical at the value of 2.4×10^{-10} mol./cm.², as is evident from figure 2. The calculated molecular area then is about 65 \AA^2 per molecule, a reasonable figure which is supported by Fischer-Hirschfelder models. If this interpretation of the results is correct, it indicates that an important advantage of the radiotracer method in conjunction with surface tension measurements is that together they may provide an insight into the composition of the surface layer.

The isotherm from the radioactivity measurements was drawn with a long slow rise after the initial steep portion because the precision of the counting procedure could not justify the presence of any break at the micelle point (70×10^{-8} mol./ml.). Represented in this fashion, the radioactivity isotherm suggests the possibility of multimolecular adsorption, and, in fact, the similarity of the observed isotherm to a Brunauer-Emmett-Teller (Brunauer 1943) isotherm for the adsorption of gas on a solid is obvious. However, the question must be examined here as to what constitutes the 'surface' in the sense of the present development. The self-absorption of ³⁵S radiation is less than 1 % for a layer of solution 10,000 Å thick. Since 1 % is at best the experimental limit of precision, in reality, therefore, an observed increase in activity could arise from any increase in concentration of molecules of surface-active agent in a layer of this depth. Consequently, no closer definition of the location of the molecules can be made. A discussion of the possibility of multilayer adsorption has been presented recently by Henniker (1949).

An alternative possibility for the slow rise in the counting isotherm is that, if the sulphosuccinate ions are associated below the critical concentration, the actual concentration of the long chain ions would be appreciably less than the analytical concentration. This would lead to a gradually increasing value of $\frac{d\gamma}{d \ln c}$ so that the surface excess based on surface tension would increase with concentration, in agreement with the radiotracer results.

Kinetics of adsorption in the interface

At high concentrations, the activity obtained for a sulphosuccinate solution was steady, but at low concentrations the counts increased with time up to a final steady value; figure 4 shows examples of each of these. These observations suggested the possibility of studying the kinetics of adsorption at the interface. Examination

indicated that although the data for each run conformed to a first order rate, the rate constant obtained was a function of the initial concentration, and the effect of concentration could not be satisfactorily taken into account.

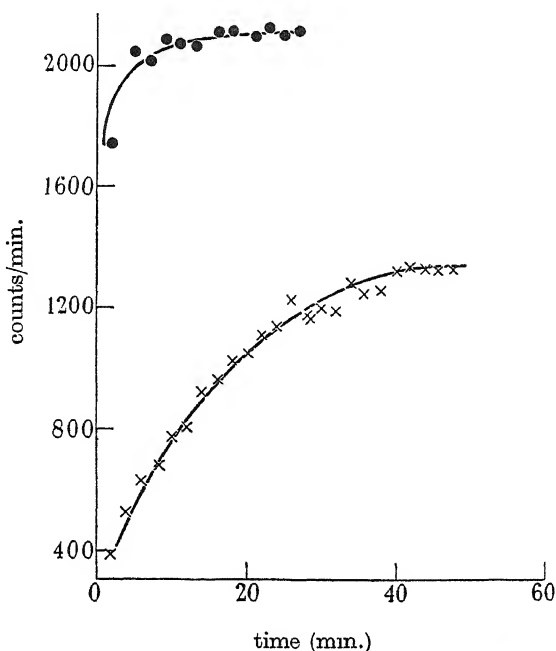


FIGURE 4. Increase in count with time of radioactive solutions ● solution 78-3, calculated count of the bulk solution = 325 counts/min.; × solution 78-1, calculated count of the bulk solution = 65 counts/min.

A treatment of the data on the basis that diffusion to the interface was rate-controlling was then made following the simple equation given by Ward & Tordai (1946; Blair 1948)

$$m_a = 2m_0 \frac{Dt^{\frac{1}{2}}}{\pi}, \quad (10)$$

where m_a = mols adsorbed at any time, t , m_0 = initial (constant) concentration, and D = diffusion coefficient. A plot of the mols adsorbed versus time on a log-log graph gave straight lines for those runs for which satisfactory adsorption/time data were obtainable. The values of the slopes should be 0.5, but this was true in only two cases, the remainder being in the neighbourhood of 0.4. The data are shown in table 2. Diffusion coefficients calculated on the basis that 0.5 was the correct exponent ranged from 0.7 to 2.2×10^{-4} cm.²/min. They varied with initial concentration, decreasing more or less regularly as the concentration increased. None the less they were of the correct order of magnitude, since it was estimated that the diffusion constant of the sulphosuccinate ion should be in the neighbourhood of 3 to 4×10^{-4} cm.²/min. (Judson 1949).

The analysis of the kinetics of adsorption has not been presented in more detail because lack of close control of the temperature, of evaporation effects, and of other

factors may have distorted the results. None the less, the radioactive technique offers a promising experimental approach, since it provides directly a measurement of the amount of material in the adsorption layer. However, it must be pointed out that interpretation is not unambiguous. This arises from the fact, as mentioned above, that the surface can only be defined within a limit of 10,000 Å so that as soon as a molecule moves into this region it should be counted. Hence, the actual step whereby the molecule becomes adsorbed in the surface may not be clearly definable.

TABLE 2. DIFFUSION COEFFICIENTS INTO AN AIR/WATER INTERFACE

initial bulk concentration, m_0 (mol./ml. $\times 10^{-8}$)	slope	$D \times 10^4$ (cm. ² /min)
1.13	0.39	0.7
0.7	0.38	1.6
0.45	0.38	0.7
0.28	0.40	2.2
0.18	0.50	1.4
0.14	0.50	1.2
0.09	0.39	1.8

APPLICATIONS

The radiotracer method outlined above seems to be applicable to surface phenomena involving adsorbed layers, either mono- or multi-molecular in nature. Although this paper describes the use of ^{35}S , it is evident that other radioisotopes having sufficiently soft radiation, such as ^{14}C , ^{45}Ca , ^{59}Fe , etc., could be employed. The discussion concerning the comparison of the surface tension and surface counting results emphasizes the desirability of being able to examine interaction in surface layers, especially that of small ions with adsorbed surface-active agents (cf. Reichenberger 1947). For example, by using labelled sulphate, the adsorption of sulphate ions in the surface layer in a solution containing both cetyl pyridinium chloride and sodium sulphate has already been measured in some preliminary work. Since the small ion can be tagged more easily than the surface-active long-chain ion, this method of studying surface adsorption and rate of formation of surface layers may be particularly useful. It may be necessary, however, to label both the long-chain ion and small gegenion in order to obtain a complete understanding of the ionic interactions in the surface layer.

The application of the method to the study of surface phenomena in systems containing several surface-active materials is under consideration. In the case of such mixtures, measurements of surface tension are not capable of yielding data that can be interpreted simply, but the tracer method will indicate the presence and amount of the tagged agent in the interface. Interaction of dissolved substances with insoluble surface layers, for example of a labelled drug or ion with a protein film, seems a possibility of great interest.

REFERENCES

- Adam, N. K. 1938 *The physics and chemistry of surfaces*. Oxford University Press.
Blair, C. M. Jr. 1948 *J. Chem. Phys.* **16**, 113.
Brady, A. P. 1949 *J. Phys. Colloid Chem.* **53**, 56.
Brunauer, S. 1943 *The adsorption of gases and vapors*. Princeton University Press.
Calvin, M., Heidelberger, C., Reid, J. C., Tolbert, B. M. & Yankwich, P. F. 1949 *Isotopic carbon*, p. 311. New York: John Wiley and Sons, Inc.
Dixon, J. K., Weith, A. J. Jr., Argyle, A. A. & Salley, D. J. 1949 *Nature* **163**, 845.
Donnan, F. G. & Barker, J. T. 1911 *Proc. Roy. Soc. A* **85**, 557.
Harkins, W. D. & Jordan, H. F. 1930 *J. Amer. Chem. Soc.* **52**, 1751.
Henniker, J. C. 1949 *Rev. Mod. Phys.* **21**, 322.
Judson, C. M. 1949 Private communication, this laboratory.
McBain, J. W. & Bolduan, O. E. A. 1943 *J. Phys. Chem.* **47**, 94.
McBain, J. W. & Dubois, R. 1929 *J. Amer. Chem. Soc.* **51**, 3534.
McBain, J. W. & Humphreys, C. W. 1932 *J. Phys. Chem.* **36**, 300.
McBain, J. W. & Swam, R. C. 1936 *Proc. Roy. Soc. A* **154**, 608.
Niederl, J. B. & Niederl, V. 1942 *Micromethods of quantitative organic analysis*, 2nd ed. chap. 7. New York: John Wiley and Sons, Inc.
Reichenberger, D. 1947 *Trans. Faraday Soc.* **43**, 467.
Ward, A. F. H. & Tordai, L. 1946 *J. Chem. Phys.* **14**, 453.
-

Diffusion and flow of gases and vapours through micropores

I. Slip flow and molecular streaming

BY P. C. CARMAN

(Communicated by E. K. Rideal, F.R.S.—Received 25 January 1950)

It has been shown that the slight minimum found by Knudsen in the plot of G against p for circular capillaries is very much exaggerated for non-circular cross-sections. Since the pore-system in a porous medium must correspond to non-circular sections, a similar minimum might be expected. Such a possibility would give rise to considerable uncertainty in applying gas permeabilities to the determination of specific surfaces of very fine powders, when flow takes place mainly by molecular streaming.

The minimum should appear when $d_p = \lambda$ (transition region). By using a number of powders of different particle size and four different gases, a very wide range of values of d_p/λ on both sides of unity has been covered. It has been established clearly that, while the plot of G against p sometimes undergoes a small and very gradual change of gradient in the transition region, no sign of a minimum has been detected. For all practical purposes, it is sufficient to assume that G against p is linear for all values of p .

For a powder of Pyrex glass microspheres, the value of $\delta/k' = 0.46$, which agrees in magnitude with theoretical expectations. For different plugs, and also for the same plug, when G against p is curved, δ/k' has different values, but the total range of variation is not large.

The foregoing conclusions are not dependent on porosity or pore texture. The pore texture of fine powders at high porosities is so non-uniform that the surface calculated from permeability is 50 % or less of its true value, but the transition from slip flow to streaming is unaffected and values of δ/k' are normal.

Flow of a gas through a porous solid is generally regarded as analogous to flow through a bundle of capillaries. If the diameter of the pores is not too small, flow-rates are inversely proportional to viscosity, in analogy to Poiseuille's law. It has also been

realized for several years (see Adzumi 1937*b*; Klinkenberg 1941; Krutter & Day 1941; and Calhoun & Yuster 1946) that, when pore diameters become comparable with or less than molecular mean free paths, viscous flow undergoes a transition to slip flow and thence to molecular streaming, just as observed by Knudsen (1909) for capillaries. Molecular streaming is really a process of diffusion. It is independent of viscosity, but obeys Graham's law, i.e. the rate under otherwise equal conditions is inversely proportional to the square root of the molecular weight. In a mixture, each gas flows independently at a rate determined by its molecular weight and *partial* pressure gradient, enabling separations to be achieved.

Fine pores correspond to a large internal surface, so that effects due to adsorption become noticeable, particularly in the case of physical adsorption of vapours. Molecules so adsorbed are more or less free to move over the solid surface, so that, if a pressure difference is maintained across a porous solid, a corresponding surface concentration gradient is produced and flow by surface diffusion can take place.

The present studies are concerned with the transition to molecular streaming and with flow by surface diffusion. To form a pore structure with controlled characteristics, it is best to compress a fine powder to a plug, in a tube of uniform bore, since the powder can be carefully selected and taken to any desired degree of compaction. Compared, say, to a porous solid such as a charcoal rod, accidental cracks and channels of relatively large dimensions are eliminated, while there is no possibility of leakage at the sides when a pressure difference is applied to produce a flow.

PROBLEMS IN TRANSITION TO MOLECULAR STREAMING

As it is not possible to deduce rigorous equations for flow in porous media, resort must be made to analogy with flow through bundles of capillaries. The first serious attempt to deduce an equation with a slip term was made by Adzumi (1937*b*), who, however, merely assumed that the porous medium was equivalent to a bundle of straight, parallel, circular capillaries. He does not seem to have been aware of the very successful treatment of Kozeny (1927) for viscous flow through porous media, and Rigden (1947) was therefore first to extend Kozeny's ideas to the slip term. Arnell (1946, 1947, 1948) independently adopted a similar approach, and a later paper by Carman & Arnell (1948) reviewed the position.

The equation arrived at may be put in the form

$$\frac{GL}{A} = \frac{p\epsilon^3}{k\eta S_0^2(1-\epsilon)^2} + \frac{8\delta\epsilon^2}{3k'(1-\epsilon)S_0} \sqrt{\left(\frac{2RT}{\pi M}\right)}. \quad (1)$$

In this, the first term on the right-hand side is the Kozeny term for viscous flow in porous media, and it has been found that $k = 5.0$. For the same porous plug and the same gas, all quantities except GL/A and p are constant, so that, plotted against one another, a straight line is obtained. From the gradient of this, assuming $k = 5$, S_0 can be calculated. The second term is the intercept at $p = 0$, so that, substituting S_0 , a value for δ/k' is obtained. It was shown by Carman & Arnell that

- (i) S_0 is in agreement with other methods of measuring particle surface,
- (ii) for a given powder, δ/k' showed no large, consistent variations with porosity,

(iii) assuming $\delta = 0.9$, nearly all values of $1/k'$ lay between 0.4 and 0.6, which was in accord with theoretical expectations.

The main application of (1) is for the determination of surface and it seems sufficient for this purpose to introduce a reasonable average value for δ/k' , e.g. 0.45, and to solve for S_0 . It is then only necessary to make a permeability test at a single value of p . This is indeed satisfactory when the slip term is less than the viscous flow term, but, as the equation is applied to finer powders, a point is eventually reached where the viscous flow term is negligible and pure molecular streaming is obtained. Now, for flow in circular capillaries, Knudsen 1909 pointed out that, when molecular mean free paths are of the same order as capillary diameter, the plot of GL/A against p is no longer linear, but passes through a minimum and then approaches the limiting value for molecular flow at low pressure. This is illustrated in figure 1, taken from

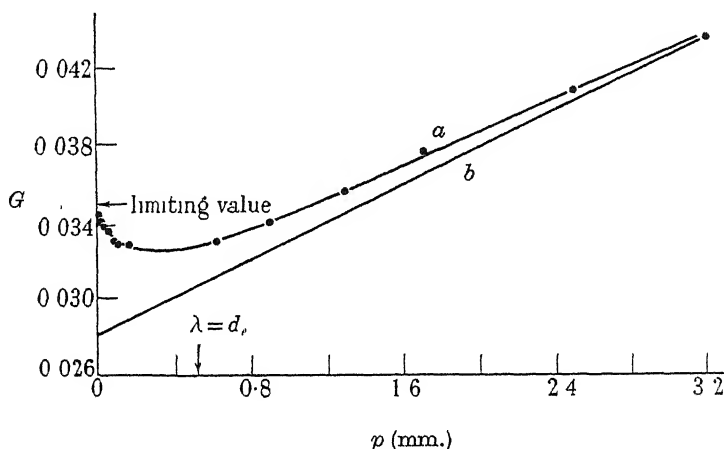


FIGURE 1. Transition region for circular capillary from Knudsen's data. *a*, observed curve; *b*, extrapolation from high pressure.

his data. If similar behaviour applies to porous media, it means that calculations of S_0 are beset by uncertainties in the 'slip term' in the very region where this is the most important term. In effect, δ/k' is no longer constant in this range, but varies toward higher values as p decreases. At first sight, it seems possible to avoid the use of the slip term by obtaining GL/A for a series of values of p , plotting one against the other, and calculating S_0 from the gradient of the straight line obtained. There are several objections to this procedure. It would only be of interest when the viscous flow term is relatively very very small or negligible, so that the gradient would be small and would be very difficult to measure accurately. Further, it is obvious from figure 1 that an apparently straight line may be obtained over a limited range of pressures, but that its gradient may be in error by several hundred percent if measurements are in the vicinity of the minimum, as would most likely be the case in examples of practical interest. To these points may be added the fact that the main feature of the permeability method has been its experimental simplicity, and its speed. It is not desirable to multiply the time taken for each determination by a larger factor.

It is clear that it is vital to make a careful study of the behaviour of the plot of GL/A against p in the 'transition region' before the permeability method can be applied with any confidence to very fine powders.

THEORY

Equations for flow in porous media can only be approached by analogy with flow in capillaries, and in particular, with non-circular capillaries. It will be shown here that such an approach leads to very disturbing conclusions.

The equation for flow in *circular* capillaries can be written

$$\frac{GL}{A} = \frac{d_e^2 p}{32\eta} + \frac{d_e p}{4\beta}, \quad (2)$$

in which the last term is the slip term. Maxwell (see Loeb 1934), and later Millikan (1923), evaluated the coefficient of external friction, β , from theoretical considerations as follows:

$$\beta = p \sqrt{\left(\frac{2M}{\pi RT}\right) \frac{f}{2-f}}, \quad (3)$$

where f is the fraction of molecules which undergo diffuse reflexion at the capillary walls. Thus

$$\frac{GL}{A} = \frac{d_e^2 p}{32\eta} + \frac{d_e}{4} \sqrt{\left(\frac{\pi RT}{2M}\right) \frac{2-f}{f}}, \quad (4)$$

and a plot of GL/A against p should be linear. The derivation of (3), however, depends on the assumption that d_e is very much greater than the mean free path, λ , where

$$\lambda = \frac{\eta}{p} \sqrt{\left(\frac{\pi RT}{2M}\right)}. \quad (5)$$

Knudsen (1909), and later Smoluchowski (1911), discussed the case where $d_e \ll \lambda$, and the outcome was equation (6) for molecular streaming:

$$\frac{GL}{A} = \frac{2d_e}{3} \left(\frac{2RT}{\pi M}\right) \frac{2-f}{f}. \quad (6)$$

This differs from the slip term in its numerical factor. Equations (4) and (6) can be combined into a general equation by introducing a numerical factor, δ , which includes the factor, $(2-f)/f$, and is a function of the ratio d_e/λ .

$$\frac{GL}{A} = \frac{d_e^2 p}{32\eta} + \frac{2\delta d_e}{3} \sqrt{\left(\frac{2RT}{\pi M}\right)}. \quad (7)$$

The relationship between δ and d_e/λ in the transition range has not proved amenable to a rigorous analysis, but theory and experiment confirm that it should approach two constant limits, namely,

$$\left. \begin{aligned} \delta &= \delta_0 = \frac{2-f_0}{f_0}, & \text{when } d_e \ll \lambda, \\ \delta &= \delta_1 = \frac{3\pi}{16} \frac{2-f_1}{f_1}, & \text{when } d_e \gg \lambda. \end{aligned} \right\} \quad (8)$$

In molecular streaming, Knudsen and subsequent workers have always found $f_0 = 1$, whence $\delta_0 = 1$. When (4) applies, f is once again constant, but appears to undergo a decrease in the transition range, with the result that δ does not decrease from 1.0 to $3\pi/16 = 0.59$, but to some intermediate value. The minimum in figure 1 arises from the variation in δ when d_e and λ are of the same order. Though Knudsen suggested an empirical function for δ in terms of d_e/λ , this is not generally applicable. In particular, when $d_e \gg \lambda$, f_1 is constant for a particular system, but varies with different gases and different surfaces, so that δ_1 shows a corresponding variation. This is illustrated in table 1. In this, values of f_1 in brackets are calculated by equation (8) from values of δ_1 obtained with capillaries; values of δ_1 in brackets are calculated from values of f_1 obtained with falling drops and concentric cylinder viscometers.

TABLE 1

system	δ_1	f_1
H ₂ —silver (a)	0.89	(0.80)
H ₂ —aluminium (a)	0.81	(0.84)
H ₂ —copper (a)	0.79	(0.85)
H ₂ —iron (a)	0.73	(0.89)
H ₂ —glass (b)	0.91	(0.79)
C ₂ H ₂ —glass (b)	0.88	(0.80)
C ₃ H ₈ —glass (b)	0.90	(0.79)
air—glass (c)	(0.73)	0.89
air—fresh shellac (c)	(0.88)	0.79
air or CO ₂ —old shellac or machined brass	(0.59)	1.0
air—mercury (c)	(0.59)	1.0
air—oil (c)	(0.73)	0.895
CO ₂ —oil (c)	(0.69)	0.92
H ₂ —oil (c)	(0.69)	0.92
He—oil (c)	(0.76)	0.90
He, H ₂ , air, O ₂ —silver oxide (d)	(0.59)	1.0

References: a, Adzumi 1939; b, Adzumi 1937a; c, Millikan 1923; d, Blankenstein 1923.

For *non-circular cross-sections*, a rigorous treatment involves considerable complexities, so that a more generally useful approach is to make use of the mean hydraulic radius, m , in place of capillary diameter, where

$$m = \frac{\text{cross-sectional area}}{\text{cross-sectional perimeter}} = \frac{\text{volume of capillary}}{\text{internal surface of capillary}},$$

and is $\frac{1}{2}d_e$ for circular sections. Then, in place of equation (2), we have

$$\frac{GL}{A} = \frac{m^2 p}{k_0 \eta} + \frac{mp}{\beta}, \quad (9)$$

where $k_0 = 2.0$ for circular sections and has only a limited range of variation for other shapes, as shown in table 2. In the slip term, β is given by (3), and, according to Millikan's derivation, the slip term in (9) should be equally valid for all shapes.

Knudsen (1909) presented a general equation for molecular streaming which can be put in the form

$$\frac{GL}{A} = \frac{8m}{3} \sqrt{\left(\frac{2RT}{\pi M}\right)}, \quad (10)$$

and which is therefore also independent of shape. Smoluchowski (1911), however, later showed that this is not correct, and equation (10) must therefore be modified to equation (11), by introducing a numerical factor, k_1 , to take care of changes of shape. Typical values calculated for k_1 are shown in table 2:

TABLE 2

shape of cross-section	k_0	k_1
circle	2.00	1.00
square	1.78	0.90
rectangle · breadth = 2 × height	1.94	0.87
breadth = 5 × height	2.38	0.78
breadth = 10 × height	2.65	0.68
parallel plates	3.00	(0)

$$\frac{GL}{A} = \frac{8m}{3k_1} \sqrt{\left(\frac{2RT}{\pi M}\right)}. \quad (11)$$

A general equation for non-circular sections thus becomes

$$\frac{GL}{A} = \frac{m^2 p}{k_0 \eta} + \frac{8m\delta}{3k_1} \sqrt{\left(\frac{2RT}{\pi M}\right)}, \quad (12)$$

in which δ has the limiting values given by equation (8) when $m \ll \lambda$ and $m \gg \lambda$ respectively. In the transition range, however, there is not only the change in δ to be taken into account, but also a variation of k_1 from the values in table 2 to a probable value of unity at high values of p .

Flow between parallel plates

The only data on the transition range for non-circular sections are those of Gaede (1913) on flow of hydrogen at 17°C between parallel plates, and it is therefore desirable to analyze them in some detail. For flow between parallel plates, separated by a distance h , $m = \frac{1}{2}h$ and $k_0 = 3.0$, provided that the ratio of the breadth of the plates, b , to the height is so large that it can be regarded as infinite. Further, $A = bh$, whence equation (12) becomes

$$G = \frac{bh^3 p}{12\eta L} + \frac{4bh^2 \delta}{3k_1 L} \sqrt{\left(\frac{2RT}{\pi M}\right)}. \quad (13)$$

Since flow took place through two parallel openings, each 1.7 cm. wide, i.e. $b = 3.4$ cm., while $L = 0.12$ cm. and the nominal value of h was 4×10^{-4} cm., the ratio of b to h was indeed very large.

TABLE 3. GAEDE'S DATA FOR PARALLEL PLATES (p IN MM.)

p	G	p	G	p	G
707	0.779	9.3	0.281	0.32	0.431
652	0.741	5.45	0.305	0.265	0.441
484	0.609	4.44	0.304	0.131	0.465
307	0.456	2.7	0.334	0.079	0.486
228	0.381	2.2	0.332	0.065	0.498
114	0.295	1.33	0.359	0.039	0.523
59	0.264	1.08	0.375	0.019	0.540
16.1	0.273	0.67	0.411		

As Gaede's analysis of his own data will be ignored, they are repeated in table 3. The transition range is shown in figure 2. The most important point, in contrast to figure 1, for circular sections, is the enormous exaggeration of the minimum, and it is proposed now to examine whether this is in accord with theoretical expectations. At high pressures, G against p gives a straight line, shown extrapolated in figure 2. The equation for this is

$$\left. \begin{aligned} G &= 8.6 \times 10^{-4}p + 0.18 & (p \text{ in mm.}) \\ &= 6.45 \times 10^{-7}p + 0.18 & (p \text{ in dynes/cm.}^2). \end{aligned} \right\} \quad (14)$$

It follows that $\frac{bh^3}{12\eta L} = 6.45 \times 10^{-7}$,

and, taking $\eta = 0.88 \times 10^{-4}$ poise for hydrogen at 17°C , the true value of h proves to be 2.88×10^{-4} cm. As Gaede himself considered direct measurement of h doubtful, and as a sufficiently wide range of high pressures was covered to establish the gradient with certainty, calculation of h from the flow data is entirely reasonable. It follows that for the intercept

$$0.18 = \frac{4bh^2\delta_1}{3k_1L} \sqrt{\left(\frac{2RT}{\pi M}\right)}.$$

In this range, k_1 is rigorously equal to unity for parallel plates, whence, substituting the value of h just obtained, we obtain

$$\delta_1 = 0.66.$$

This compares reasonably with values given in table 1. The pressure at which $\lambda = h$ is 31 mm., which is in the vicinity of the minimum in figure 2.

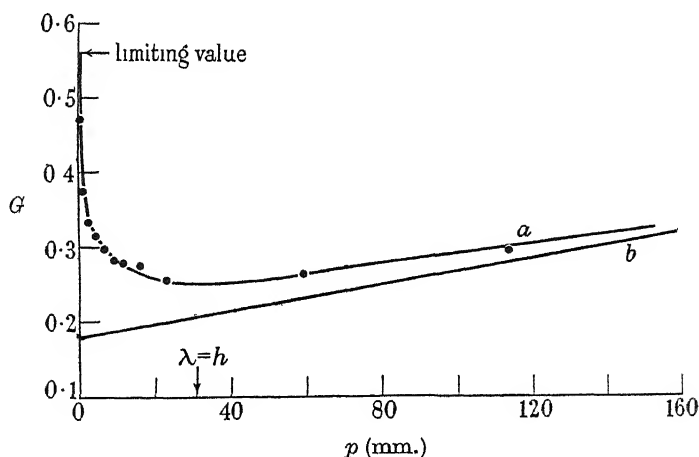


FIGURE 2. Transition region for a slit, using Gaede's data. *a*, observed curve; *b*, extrapolation from high pressures.

At low pressures, a constant, limiting value of G was approached very slowly, compared with circular capillaries. The apparent limit obtained on extrapolating the data at the lowest pressures is 0.56, but it is possible that even this was too low. The

reason for this is apparent if we consider that pure molecular streaming requires λ to be very much larger than *all* dimensions of a cross-section. The transition to molecular flow starts when $\lambda \simeq h$, but it cannot be complete till $\lambda \gg b$. For molecular streaming, we have seen that we can always take $\delta = 1$, whence, putting

$$0.56 = \frac{4bh^2}{3k_1 L} \sqrt{\left(\frac{2RT}{\pi M}\right)},$$

we obtain $k_1 = 0.49$. Now, the theoretical value of k_1 for molecular streaming between parallel plates is zero, but this imposes the impossible condition that both L/h and b/h are infinite. In an analysis of Gaede's data, Klose (1931) applied Smoluchowski's formula for a rectangular slit, which gives $k_1 \simeq 0.25$, but this formula assumes that $L \gg b$, $L \gg h$, which is not true in this case. The correct formula for a rectangular slit when $b \gg L \gg h$, according to Clausing (1929), is

$$k_1 = \frac{8}{3} \frac{1}{\ln(L/h)},$$

which gives $k_1 = 0.44$. Bearing in mind that the experimental limiting value of G is likely to be too low, and hence the value of k_1 too high, the agreement between calculated and experimental values of k_1 is excellent.

We can thus conclude that equation (12) would apply equally well to other cross-sections, using appropriate values of k_0 and k_1 .

Porous media

The approach used by Kozeny (1927) and extended by Rigden (1947) and Arnell (1946, 1947, 1948) is to assume that the porous medium is equivalent to a collection of non-circular capillaries, and to substitute in equation (12). The mean hydraulic radius in this case is given by

$$m = \frac{\epsilon}{S_0(1-\epsilon)}, \quad (15)$$

and correction has to be made for the facts that flow is restricted to a fraction, ϵ , of the total cross-section, and that a tortuous path of actual length, L_e , is traversed through a plug of length L . Thus (12) becomes

$$\frac{GL}{A\epsilon} \left(\frac{L_e}{L}\right)^2 = \frac{m^2 p}{k_0 \eta} + \frac{8m\delta}{3k_1} \sqrt{\left(\frac{2RT}{\pi M}\right)}. \quad (16)$$

Substituting for m , and putting

$$k = k_0 \left(\frac{L_e}{L}\right)^2 \quad \text{and} \quad k' = k_1 \left(\frac{L_e}{L}\right)^2 = \frac{k_1 k}{k_0}, \quad (17)$$

we obtain equation (1).

Suppose, now, that we consider possible values of δ/k' , i.e. $\delta k_0/k_1 k$ in equation (1). For the mean pore diameter, d_e , we can write

$$d_e = 4m = \frac{4\epsilon}{S_0(1-\epsilon)}. \quad (18)$$

Then, if $d_e \gg \lambda$, we can take $k_1 = 1$, while δ has a lower limit of 0.59, but can generally be expected to have somewhat larger values up to about 0.9. Taking the average shape as more or less rectangular, a typical value of k_0 should be 2.5, while $k = 5.0$. Thus, δ/k' would generally fall in the range 0.3 to 0.5, which accords with the data reviewed by Carman & Arnell (1948).

The possibility of changes in δ/k' in the transition range, however, seems very probable. We have seen that, for circular capillaries, changes in δ in this range seldom exceed 10 to 20 %; but, for parallel plates, a change of over 300 % in δ/k' was observed. This is admittedly an extreme case. Suppose we consider the type of change in δ/k' which might be expected in the transition range of a porous plug. If, say, we take $k_1 = 1$, $k_0 = 2.5$, $\delta_1 = 0.8$ at high pressures, the calculated value of the ratio is 0.4. For molecular streaming, $\delta = 1$, and a rectangular slit which gives $k_0 = 2.5$ should also give $k_1 \simeq 0.7$, whence $\delta/k' \simeq 0.7$. For determination of specific surface of very fine powders such variations would be serious, since it must be remembered that δ/k' is actually measured empirically, and this is only possible above the transition range. It is shown in the present work that δ/k' shows remarkably little variation in the transition range; but this could not have been predicted from theoretical considerations.

EXPERIMENTAL

The apparatus was intended primarily for the surface-flow measurements to be described in part II (Carman & Malherbe 1950). For these, a stationary flow-rate must be maintained for long periods to produce equilibrium. A diagram of the apparatus is shown in figure 3. The simplest method for maintaining constant pressures was to maintain a slow leak of the gas being used through the two mercury-filled columns, *A* and *B*. The heights in these were adjusted by running mercury in at taps, T_3 and T_7 , or running it out at the bottom. To begin an experiment the gas supply was connected at T_1 . Taps T_1 , T_3 , T_7 and T_{12} were closed, while all others were opened, and the apparatus was pumped out (T_4 was only opened when pumping was taken to a very hard vacuum). With T_4 , T_{10} , T_{11} closed, T_1 was opened to allow gas to enter. Then, with T_2 and T_5 closed, T_{10} was opened till a stream of gas was bubbling through *A* and *B*, after which T_{10} was adjusted till a slow, steady rate of bubbling was registered in *C* by a layer of dibutyl phthalate just covering the tip of the inlet tube. The pressure, p_1 , was then recorded by manometer, M_1 , and the pressure drop, Δp , by M_2 . To stabilize pressures, a surge volume, *D*, was needed between *A* and *B*. In some cases, for small values of Δp , dibutyl phthalate was used in *B* and a manometer filled with the same liquid (not shown in the diagram) replaced M_2 . Both M_1 and M_2 were constructed of 14 mm. o.d. tubing to avoid meniscus errors. By careful attention to the design of *A* and *B*, pressures could be maintained to 0.1 mm. for long periods.

Before starting the experiment, the plug was formed in the permeability tube, *E*, and one end was connected to a short piece of pressure tubing, *F*. The other end was tapered to fit into the A_{10} joint, *G*. For accurate measurements of flow-rates, which were sometimes as low as 1 ml./hr., it was decided that only direct measurement of volume at constant pressure could be used. This was done with the burette, *H*, connected at one end to an A_7 joint, and, at the other, by pressure tubing to the

mercury reservoir, *J*. For, *H*, a 10 ml. and a 1 ml. burette were used interchangeably. To maintain constant pressure, the method of Kohman (1929) was used, namely, generation of gases between platinum electrodes in an oxalic acid solution in the electrolytic cell, *K*. With T_{13} closed and T_{12} open, this displaces mercury from *J* to *H* at the same rate as gas is withdrawn from *H*, thereby keeping the pressure in *H* constant at p_1 . When T_8 is closed, a volume of gas at p_1 is trapped between the plug in *E*, mercury in *H* and differential manometers, M_3 and M_4 , filled with mercury and dibutyl phthalate, respectively. As gas passes through the plug, mercury is drawn away from the upper platinum contact in M_3 . This activates a relay system which brings *K* into operation, compressing the gas in *H* until the contact in *M* is remade. The pressure in *H* is thus automatically kept at p_1 . The stopper, *L*, is used for filling

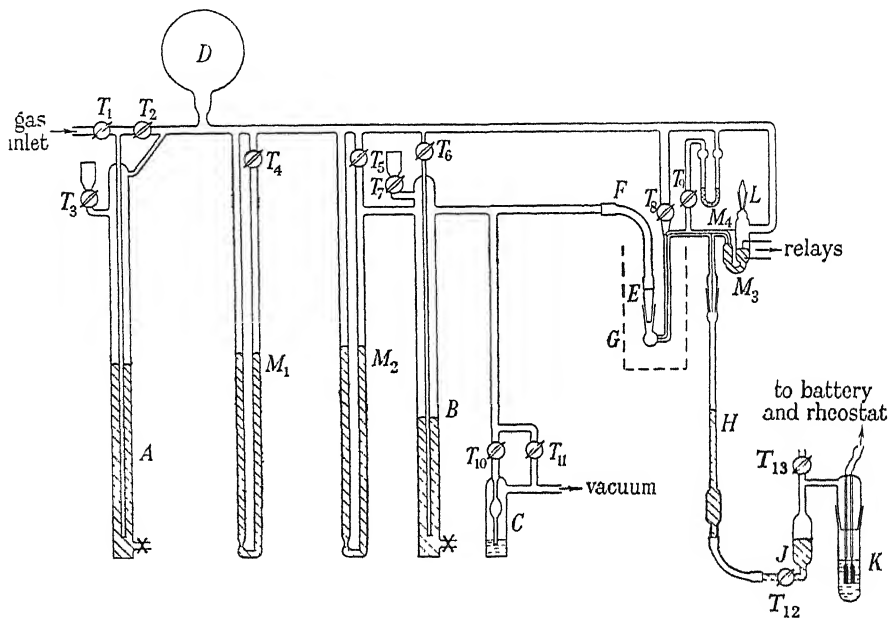


FIGURE 3

or cleaning M_3 . The manometer, M_4 , is an auxiliary, used only when M_3 is being filled, since it is very much more sensitive, and only with its use can exactly the right amount of mercury be placed in M_3 to ensure that pressures on each side of M_3 are identical when the contact is functioning. During an experiment, tap T_9 is kept closed. When *H* is full, tap T_8 is opened, and T_{13} is cracked open, allowing gas to escape from *J* till a new charge of gas has been drawn into *H*. When T_8 and T_{13} are closed, a new measurement begins. It should be noted that repeated refilling of *H* involves no disturbance of the steady flow-rate through the plug in *E*. Successful operation depends on the sensitivity of M_3 . The trapped volume must be kept as small as possible, so that capillaries are used as shown; and, since a wide tube is desirable for manometer, the contacts are placed at the surface away from *H* since they would take considerable space. The mercury surface must remain scrupulously clean, and must not become oxidized by sparking. As the relay operated at $50 \mu\text{A}$, this gave no trouble. For the cell, *K*, a 12 V battery was used, with a rheostat, so that

the rate of electrolysis in each experiment could be adjusted until it approximately balances the rate of flow through the plug. This was necessary if M_3 was to function under optimum conditions.

The lower the pressure, p_1 , the less sensitive will be M_3 . In practice, no experiments were carried out with $p_1 < 25$ mm. It is thus evident that, compared with the constant-volume type of apparatus used by Knudsen and others, the present apparatus is less simple and is restricted to a very limited range of mean pressures, roughly, from 15 to 650 mm. (atmospheric pressure at Pretoria, 660 mm.). For the work in the present paper, it is admittedly much less suitable than a constant-volume apparatus, but, within its limits, it was capable of giving accurate results, and it is not felt that it has affected the conclusions reached.

Usually, in making a series of runs at different pressures, Δp was left unchanged, and only p_1 was altered by varying the level of mercury in A . Many experiments, however, have confirmed that the same value of $p = \frac{1}{2}(p_1 + p_2)$ always give the same value of $G = p_1/\Delta p t$, where t equals time in seconds for 1 ml. at pressure, p_1 , to flow through the plug, irrespective of the distribution of p_1 and p_2 .

It will be noted that the design allows E to be surrounded by a constant-temperature bath, and this was always used. No control of temperature in H was used. If temperatures in E and H are, respectively, T and $T_1^\circ\text{K}$, then 1 ml. in H corresponds to T/T_1 ml. in E . The time for flow of 1 ml. at $T^\circ\text{K}$ must therefore be corrected by the formula

$$t_{\text{corr}} = t \times \frac{T_1}{T}.$$

Formation of a plug of known dimensions is shown in figure 3a. The tube, E , is of polished, hardened steel with an accurate bore of 0.705 cm. The weight of the plug is found by difference, and its length by compressing between hardened steel plungers of total length 10 cm. and diameter 0.7 cm. The increase over 10 cm. is the length, L , of the plug.

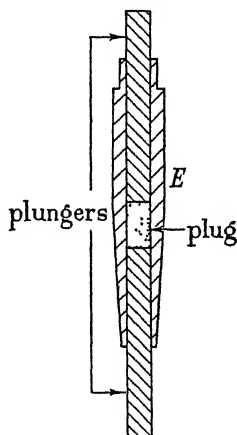


FIGURE 3a. Formation of plugs in permeability tube, E .

Owing to the limited range of pressures which could be covered for a single gas, experiments were carried out with four gases, covering a wide range of molecular weights, namely H_2 (2.016), air (28.8), CO_2 (44) and CF_2Cl_2 (121). The first was

generated from zinc and sulphuric acid and purified in the usual way; air was freed from CO_2 and water vapour; CO_2 was taken from a cylinder and dried. It was tested for purity by absorption in aqueous potash. Freon-12, CF_2Cl_2 , was also taken from a cylinder, the middle portion of the contents of each full cylinder only being used.

Measurements were made at 20°C , and viscosities were determined by sealing a 2 cm. length of capillary with a nominal bore of 0.05 mm. to an A_{10} joint and fitting this in place of E . Measurements were made at different pressures and data were treated in the same way as those for porous plugs, i.e. by plotting G against p , a straight line was obtained, for which the gradient is inversely proportional to viscosity and the intercept to the inverse square root of the molecular weight. It was assumed that η for air was 1.82×10^{-4} poise and other viscosities were calculated from this. The viscosities of H_2 and CO_2 agreed with the literature, and that of CF_2Cl_2 with Benning & Markwood (1939). As a check on accuracy, intercepts were multiplied by \sqrt{M} and found to give reasonably constant products. The values accepted in the following were:

gas	viscosity at 20°C (poise)
H_2	0.89×10^{-4}
air	1.82×10^{-4}
CO_2	1.48×10^{-4}
CF_2Cl_2	1.24×10^{-4}

RESULTS

For the following work, equation (1) may be written as

$$\frac{GL}{A} = \frac{F_1 p}{5\eta S_0^2} + \frac{2.13\delta F_2}{k'S_0} \sqrt{\left(\frac{RT}{M}\right)}, \quad (1a)$$

i.e. $k = 5.0$, $\frac{8}{3}\sqrt{(2/\pi)} = 2.13$, and F_1 and F_2 are the porosity functions, $\epsilon_3/(1-\epsilon)^2$ and $\epsilon^2/(1-\epsilon)$. The value of A was always 0.39 cm.

At least at pressures well above the transition range, G against p should give a straight line

$$G = ap + b.$$

Then, from equation (1a), we can calculate S_0 from the gradient

$$S_0 = \sqrt{\left(\frac{F_1 A}{5a\eta L}\right)}.$$

This is in turn used to calculate δ/k' from the intercept

$$\frac{\delta}{k'} = \frac{S_0 b L}{2.13 F_2 A} \sqrt{\left(\frac{M}{RT}\right)}.$$

The first experiments were carried out with a relatively coarse powder so that the slip term should be relatively small. Since a clean sample of Transvaal chromite was available, crushed and carefully air-separated into size fractions, the 5 to 10μ fraction was selected. Experimental points for G against p fell closely on a straight line for each gas, and in table 4, gradients and intercepts are given.

Though the products $a\eta$ are fairly constant, the variations were more than expected from experimental errors, and showed a definite trend with molecular weight.

Similar variations applied to the products, $b\sqrt{M}$, though, as the intercepts were relatively small in this case, the variations could have been attributed to experimental errors. It was further noted that experimental points for each gas seemed sometimes to lie on a line with a slight curvature. Since the deviations from linearity could not be tested by extending the range of pressures, it was decided to plot data for all four gases on a single graph. If, in equation (1a), $1/\lambda$ is replaced for p , using equation (5), we obtain

$$\frac{GL}{A} \sqrt{\left(\frac{M}{RT}\right)} = \frac{F_1}{5\lambda S_0^2} \sqrt{\frac{\pi}{2}} + \frac{2 \cdot 13 \delta F_2}{k' S_0}.$$

Thus, since temperatures are constant, $G\sqrt{M}$ against $1/\lambda$ should give a straight line. Results have been so plotted in figure 4 and it can be seen at once that the curvature

TABLE 4. EXPERIMENTS WITH CRUSHED CHROMITE FRACTION

$L = 3.66$, $\rho = 4.37$, $\epsilon = 0.405$, $F_1 = 0.188$, $F_2 = 0.276$

gas	a (p in mm.) $10^{-4} \times$	$a\eta$ $10^{-8} \times$	b	$b\sqrt{M}$
H ₂	9.00	8.01	0.330	0.468
air	4.29	7.82	0.090	0.484
CO ₂	5.16	7.65	0.072	0.476
CF ₂ Cl ₂	6.06	7.52	0.046	0.506

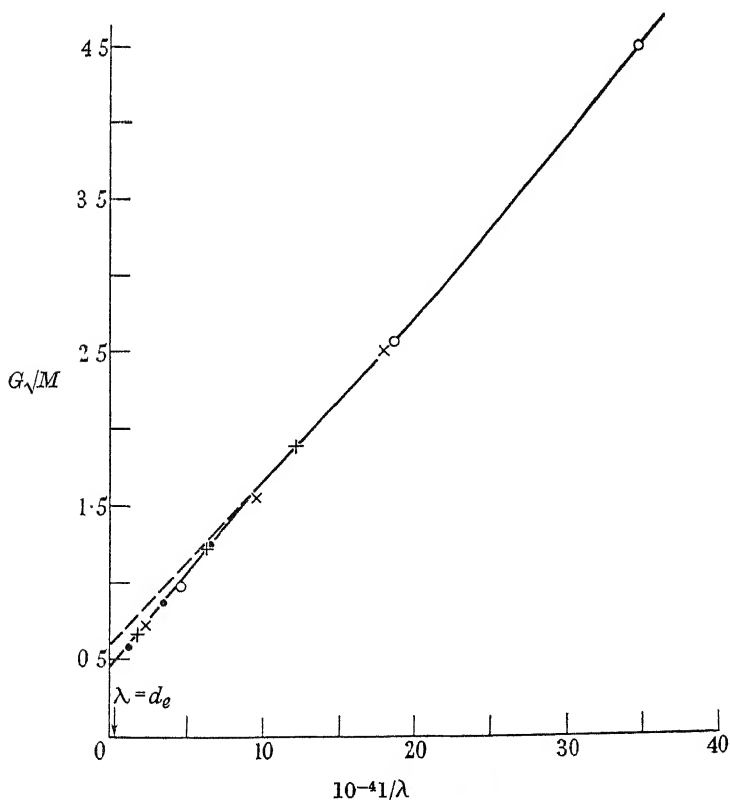


FIGURE 4. Correlation of data for 5 to 10 μ chromite. ●, H₂; +, air; ×, CO₂; ○, CF₂Cl₂.

is real, since the various gases overlap, and points at the lower values of $1/\lambda$ for one gas interpolate exactly between the higher values of $1/\lambda$ for the previous gas. At high values of $1/\lambda$, the curve approaches linearity, and, from the gradient and intercept of this line, $S_0 = 8560 \text{ cm.}^{-1}$ and $\delta/k' = 0.526$. Using this value of S_0 , the value of $1/\lambda$ for $\lambda = d_e$ is 0.31×10^4 , and this value is marked on figure 4. The deviation from linearity becomes detectable when $d_e/\lambda < 20$, and corresponds to a decrease in δ/k' . If it is assumed that the curve becomes linear again at low value of d_e/λ , and extrapolation is made to $1/\lambda = 0$, the intercept so obtained gives $\delta/k' = 0.394$, assuming $S_0 = 8560 \text{ cm.}^{-1}$.

Next considered was a Pyrex glass powder consisting of perfectly spherical particles, prepared by the method of Bloomquist & Clark (1940), and separated by sedimentation into size fractions. The finest fraction of sufficient weight to form

TABLE 5. EXPERIMENTS WITH FINE GLASS SPHERES

$L=3.06, \quad \rho=2.19, \quad \epsilon=0.435, \quad F_1=0.258, \quad F_2=0.335$						
	a					
gas	(p in mm.)		$a\eta$		$S_0 (\text{cm.}^{-1})$	δ/k'
	$10^{-4} \times$	b	$10^{-8} \times$	$b\sqrt{M}$	$10^4 \times$	
H ₂	2.10	0.215	1.87	0.305	2.15	0.456
air	1.04	0.058	1.89	0.312	2.14	0.467
CO ₂	1.29	0.046	1.91	0.305	2.14	0.456
CF ₂ Cl ₂	1.56	0.028	1.93	0.308	2.12	0.461

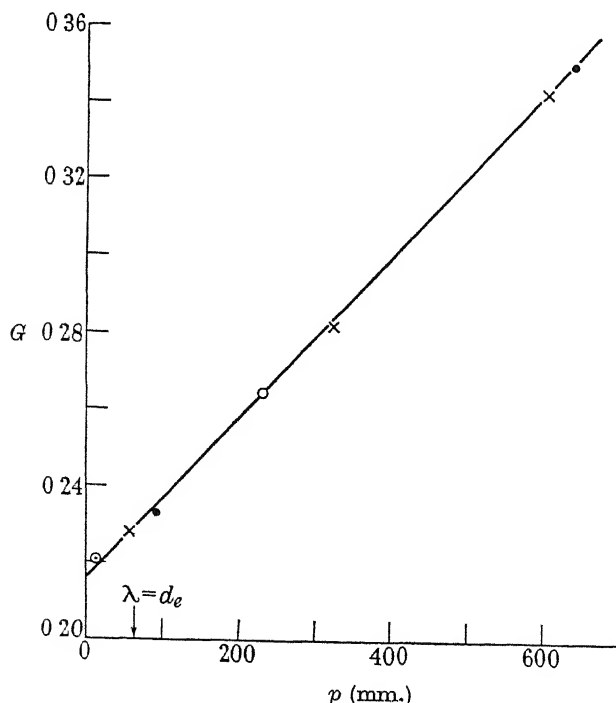


FIGURE 5. Flow of hydrogen through 2 to 5 μ glass microspheres.
Values of Δp ; ○, 18 mm.; ●, 28 mm.; ×, 96 mm.

a plug was 2 to 5μ . Microscopic examination showed this to have a surface mean diameter of 2.89μ . Data obtained in flow measurements are summarized in table 5. In this case, perfectly linear plots of G against p were obtained for each gas, and variations of $a\eta$ and $b\sqrt{M}$ can be reasonably attributed to experimental error. Values of δ/k' in the table were calculated from the mean value of S_0 . The latter corresponds to a surface mean diameter of 2.80μ , and thus agrees with the microscopic diameter. From the value of S_0 , we find $\lambda = d_e$ when $p = 64\text{ mm.}$ for H_2 , and the plot of G against p for this gas is given in figure 5. It is linear within experimental error down to $p = 12.3\text{ mm.}$, i.e. to $d_e/\lambda < 0.2$.

To follow the behaviour at lower values of d_e/λ , a finer powder is needed. A polishing powder marketed by Linde Air Products Co. consists entirely of α -alumina of over 99.9 % purity, and of very uniform particle size, approximately 0.3μ . The surface by nitrogen adsorption was $1.65 \times 10^5\text{ cm.}^{-1}$, which agrees excellently with the permeability surface given by table 6 and it therefore seemed ideally suited for the present studies.

TABLE 6. EXPERIMENTS WITH $\alpha\text{-Al}_2\text{O}_3$ POLISHING POWDER

$L = 1.11, \quad \rho = 4.04, \quad \epsilon = 0.475, \quad F_1 = 0.390, \quad F_2 = 0.430$						
a						
gas	$(p \text{ in mm.})$ $10^{-5} \times$	b	$a\eta$ $10^{-9} \times$	$b\sqrt{M}$	$S_0(\text{cm.}^{-1})$ $10^5 \times$	δ/k'
H_2	1.42	0.0953	1.26	0.135	1.71	0.464
air	0.68	0.0253	1.24	0.135	1.71	0.464
CO_2	0.80	0.0202	1.19	0.134	1.75	0.462
CF_2Cl_2	0.99	0.0120	1.23	0.132	1.72	0.455

Table 6 summarizes data on a plug compressed to $\epsilon = 0.475$. Plots of G against p were all linear, and values of $b\sqrt{M}$ were in excellent agreement. Some variation in $a\eta$ is to be expected, for viscous flow plays a smaller part than in previous tables and experimental values are therefore more open to error. Of most significance are the experiments with hydrogen, since $\lambda = d_e$ when $p = 435\text{ mm.}$, and ratios of d_e/λ could thus be followed down to 0.038. The data are plotted in figure 6, and provide strong evidence that linearity is maintained down to $p = 0$. There is no sign of any minimum similar to figures 1 and 2.

Summarizing, it appears that, at high values of d_e/λ , G against p always gives a straight line. When d_e/λ decreased towards unity, a departure from linearity is sometimes observed, as in figure 4, but no marked change of gradient is ever experienced. When d_e/λ approaches zero, G against p is linear, whether there is a change of gradient in the transition region or not.

Thus far, only plugs with porosities under $\epsilon = 0.5$ have been considered. It is of interest to inquire whether the nature of the pore system affects the conclusions just reached. The problem does not arise for the relatively coarse powders in tables 4 and 5, since no other porosities can be obtained; but very fine powders tend to give high porosities, owing to surface forces causing the particles to form more or less tightly bound aggregates. Thus, the porosity of the plug in table 6 could only be obtained after powerful compression of the plug between hardened steel plungers in a vice.

A high porosity thus corresponds to voids of relatively large size between aggregates in a plug, and the resulting non-uniformity of pore size gives a permeability which is larger than would be expected for the same average porosity if pore sizes were uniform. Consequently, values of S_0 calculated by equation (1a) are too low. Compression of the plug tends to close up large voids first, thereby giving a more uniform texture. Thus for the $\alpha\text{-Al}_2\text{O}_3$ powder, values of S_0 do not change much below $\epsilon = 0.60$, showing that the pore structure is fairly uniform in this range, but they decrease rapidly when $\epsilon > 0.65$.

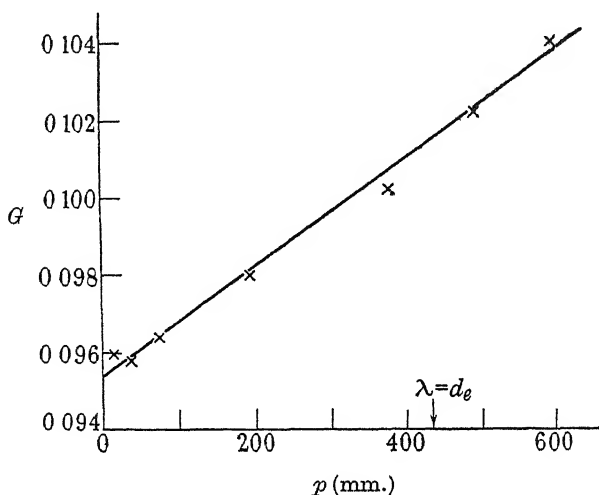


FIGURE 6. Flow of hydrogen through Al_2O_3 . $\epsilon = 0.475$.

Plugs were therefore formed from the same powder with $\epsilon = 0.636$ and $\epsilon = 0.780$, respectively. Data for the former are summarized in table 7. As expected from the relatively small change in S_0 , the pore system is still uniform, and the same linear relationships and agreement between various gases is observed as in table 6. For $\epsilon = 0.780$, similar relationships to those in table 4 were obtained, and data are therefore plotted as $G\sqrt{M}$ against $1/\lambda$ in figure 7, which is obviously closely similar to figure 4. Calculating S_0 from the gradient at high values of $1/\lambda$, we obtain $S_0 = 0.865 \times 10^5 \text{ cm.}^{-1}$ and the intercept from extrapolating this line gives $\delta/k' = 0.57$. The low value of S_0 shows that the pore system is highly non-uniform, yet the general correlation of data for different gases is exactly similar to figure 4, which corresponds to a plug with a low porosity and, presumably, a uniform pore system. In fact, the main effect of the high porosity has been to make the range of d_e/λ values covered by this fine powder similar to that covered by the coarse powder in table 4 at a low porosity. It seems reasonable to conclude that, in these plugs, and, indeed, in any porous structure, the degree of uniformity of the pore system plays no part in the transition from slip flow to molecular streaming, the only important variable being the ratio d_e/λ . On extrapolating figure 7 to $1/\lambda = 0$, an intercept is obtained which gives $\delta/k' = 0.488$.

To test the foregoing conclusion for a lower range of values of d_e/λ , experiments were carried out with a finer powder. A calcium carbonate powder with a surface of

$6.52 \times 10^5 \text{ cm.}^{-1}$ by nitrogen adsorption was chosen. Compressed to $\epsilon = 0.45$ the permeability surface was $7.7 \times 10^5 \text{ cm.}^{-1}$, showing uniform pore structure. At a porosity, $\epsilon = 0.720$, an average value of $S_0 = 3.67 \times 10^5 \text{ cm.}^{-1}$ was obtained, indicating a non-uniform pore structure. Owing to the high porosity, the range

TABLE 7. EXPERIMENTS WITH $\alpha = \text{Al}_2\text{O}_3$ POLISHING POWDER

$L = 2.65$, $\rho = 4.04$, $\epsilon = 0.636$, $F_1 = 1.94$, $F_2 = 1.11$

gas	α (p in mm.) $10^{-5} \times$	b	$a\eta$ $10^{-9} \times$	$b\sqrt{M}$	$S_0 (\text{cm.}^{-1})$ $10^5 \times$	δ/k'
H ₂	3.43	0.1175	3.05	0.167	1.58	0.498
air	1.67	0.0310	3.04	0.166	1.58	0.495
CO ₂	1.88	0.0247	2.78	0.164	1.66	0.490
CF ₂ Cl ₂	2.26	0.0149	2.80	0.164	1.65	0.490

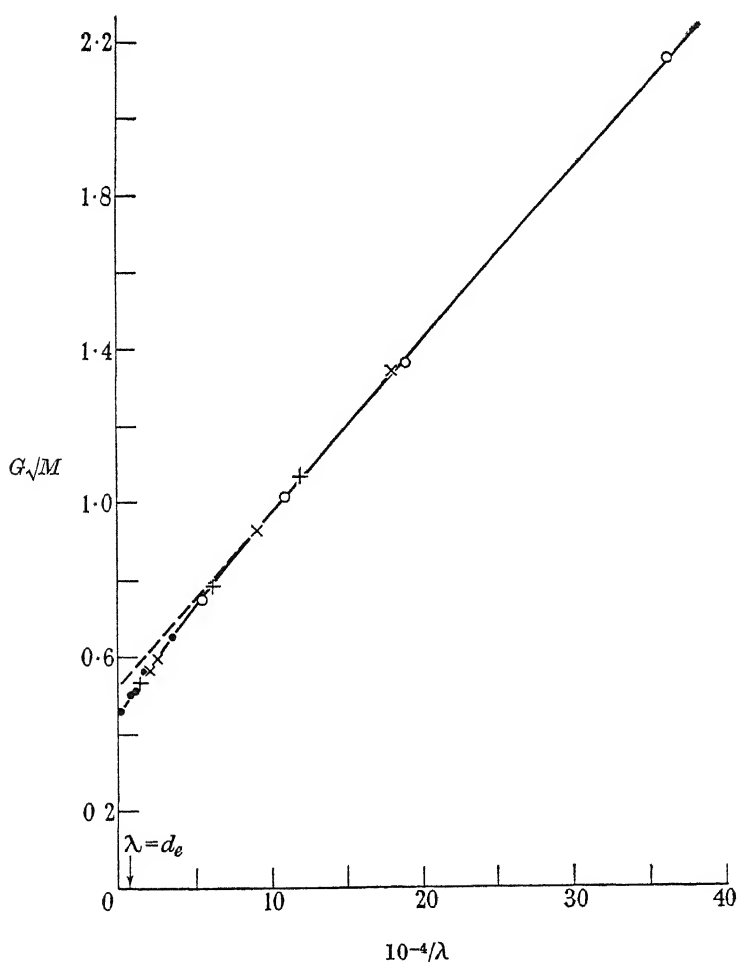


FIGURE 7. Correlation of data for $\alpha\text{-Al}_2\text{O}_3$ plug. $\epsilon = 0.78$.
 ●, H₂; +, air; ×, CO₂; ○, CF₂Cl₂.

of d_e/λ values covered was similar to that in table 6, and it is obvious that the remarks applied to table 6 apply equally well to the data in table 8, confirming that non-uniformity of pore structure has neither affected the correlation between various gases nor produced any unusual values of δ/k' .

TABLE 8. EXPERIMENTS WITH FINE CALCIUM CARBONATE

$$L=1.60, \quad \rho=2.65, \quad \epsilon=0.720, \quad F_1=4.76, \quad F_2=1.85.$$

gas	a (p in mm.) $10^{-5} \times$	b	$a\eta$ $10^{-9} \times$	$b\sqrt{M}$	S_0 (cm. ⁻¹) $10^5 \times$	δ/k'
H ₂	2.58	0.1305	2.30	0.185	3.67	0.452
air	1.28	0.0345	2.33	0.185	3.64	0.452
CO ₂	1.55	0.0268	2.29	0.178	3.67	0.435
CF ₂ Cl ₂	1.80	0.0166	2.23	0.183	3.72	0.448

One further set of experiments was carried out with a single plug of a silica powder, also furnished by Linde Air Products Co., which was so fine that $d_e \ll \lambda$ in all experiments. The surface by nitrogen adsorption was $S_0 = 4.9 \times 10^6$ cm.⁻¹, and the mean value calculated from the permeability of the plug under consideration was $S_0 = 2.45 \times 10^6$ cm.⁻¹. More highly compacted plugs were not tested so extensively, but gave values of S_0 much nearer to the adsorption surface. In the experiments, various gases at a number of pressures and temperatures were observed, and it will be clear from equation (1a) that, if the viscous flow term can be neglected, $G\sqrt{(M/T)}$ should give a constant value, provided δ/k' does not show large variations for different gases. Table 9 bears out the previous studies that such is not the case. It will be noted that CF₂Cl₂ and, to a greater extent, C₂F₄Cl₂, at room temperature, give high values. The normal boiling points of these are, respectively, -28 and 3.8°C, so that, on this fine powder, physical adsorption takes place to a degree sufficient for surface flow to be observed. Discussion of this belongs to the next paper. At 50°C, as would be expected, this effect has become negligible.

TABLE 9. EXPERIMENTS WITH LINDE SILICA PLUG

$$L=0.84, \quad \rho=2.2, \quad \epsilon=0.667, \quad F_1=2.67, \quad F_2=1.33$$

gas	temperature (°C)	M	p (mm.)	$G\sqrt{(M/T)}$ $10^{-3} \times$
H ₂	24	2.02	388	2.16
	24		584	2.20
	50		584	2.24
air	-21.5	28.8	419	2.16
	24		118	2.20
	24		394	2.23
	50		445	2.24
CO ₂	24	44	363	2.29
	50		409	2.21
CF ₂ Cl ₂	24	121	115	2.37
	24		358	2.44
	50		391	2.26
C ₂ F ₄ Cl ₂	24	171	357	2.55
	50		463	2.27

Note added in proof, 22 June 1950. Since completing the foregoing paper, the attention of the writer has been drawn to a theoretical paper by W. G. Pollard & R. D. Present (*Phys. Rev.* 1948, **73**, 762), in which it is deduced that the relation between G and p should remain linear down to $p = 0$ for porous media. They point out that the increase in G at low pressures which produces the minimum observed by Knudsen and others for straight tubes is due to an increasing proportion of long molecular paths between successive wall collisions, due to the fact that $\lambda > d_e$ but still $< \frac{1}{2}L$. When $\lambda > \frac{1}{2}L$, a limiting rate of molecular flow is approached which is less than that for a tube of infinite length. In a porous bed, which corresponds to a kinked tube, no straight path can much exceed d_e without collision, hence no increase in rate and no minimum is observed. According to the writer, their arguments do not require *exact* linearity. A small change of gradient in *either* direction in the transition region, such as shown by certain plugs in the present paper, should be a normal expectation.

A relatively minor point made by the same authors is that equation (10) of the present paper was incorrectly given by Knudsen and should be

$$\frac{GL}{A} = 2m \sqrt{\left(\frac{\pi RT}{2M}\right)}.$$

This makes no difference to the treatment in the present paper.

SYMBOLS

(c.g.s. units are used except where specifically stated otherwise)

- L = length of capillary or porous plug.
- A = cross-sectional area of capillary or porous plug.
- M = molecular weight.
- R = gas constant = 8.31×10^7 .
- η = gas viscosity at plug temperature.
- λ = molecular mean free path at plug temperature and mean pressure, p .
- β = coefficient of external friction.
- f = fraction of diffuse reflexions.
- δ = parameter in slip term, defined in equations (7) and (8).
- k_0 = shape in general equations for viscous flow through capillaries.
- k_1 = shape factor in general equation for molecular streaming through capillaries.
- k = Kozeny constant = 5.0.
- k' = corresponding constant in slip term for porous media.
- a and b = gradient and intercept in $G = ap + b$.
- $\Delta p = (p_1 - p_2)$ = pressure difference across plug. (N.B. Pressures are recorded as mm. of mercury, but are converted to dynes/cm.² in all calculations).
- p = $\frac{1}{2}(p_1 + p_2)$ = mean pressure.
- t = time in seconds for 1 ml. gas at pressure p and temperature T to flow through plug or capillary.
- T = absolute temperature of plug.
- T_1 = absolute temperature in gas burette.

$G = p_1/\Delta pt.$

ϵ = porosity.

F_1 and F_2 = porosity functions, $\frac{\epsilon^2}{(1-\epsilon)^2}$ and $\frac{\epsilon^3}{(1-\epsilon)}$.

S_0 = specific surface in $\text{cm.}^2/\text{cm.}^3$ or cm.^{-1} .

m = mean hydraulic radius = $\frac{\epsilon}{S_0(1-\epsilon)}$ for porous media.

d_e = diameter of circular capillary and = $4m = \frac{4\epsilon}{S_0(1-\epsilon)}$ for porous media.

The writer wishes to thank the South African Council for Scientific and Industrial Research for permission to publish this paper, and also to acknowledge gifts of the following powders: B. W. Wilson, Australian C.S.I.R.O. for glass microspheres; Witco Chemical Co., N.Y., for Witcarb-R calcium carbonate; Linde Air Products Co., N.Y., for Linde silica.

REFERENCES

- Adzumi, H. 1937a *Bull. Chem. Soc. Japan*, **12**, 292.
 Adzumi, H. 1937b *Bull. Chem. Soc. Japan*, **12**, 304.
 Adzumi, H. 1939 *Bull. Chem. Soc. Japan*, **14**, 343.
 Arnell, J. C. 1946 *Canad. J. Res.* **24** A, 103, *Canad. J. Res.* 1947, **25** A, 191; *Canad. J. Res.* 1948, **26** A, 29.
 Benning, A. F. & Markwood, W. H. 1939 *Refrig. Engng*, p. 243.
 Blankenstein, E. 1923 *Phys. Rev.* **22**, 582.
 Bloomquist, C. R. & Clark, A. 1940 *Industr. Engng Chem.* (Anal. ed.), **12**, 61.
 Calhoun, J. C. & Yuster, S. T. 1946 *Drilling and production practice*, p. 335.
 Carman, P. C. 1938 *J. Soc. Chem. Ind., Lond.*, **57**, 225; *J. Soc. Chem. Ind., Lond.*, 1939, **58**, 1.
 Carman, P. C. & Arnell, J. C. 1948 *Canad. J. Res.* **26** A, 129.
 Carman, P. C. & Malherbe, P. le R. 1950 *Proc. Roy. Soc. A.* **203**, 165.
 Clausius, P. 1929 *Physica*, **9**, 65.
 Gaede, W. 1913 *Ann. Phys., Lpz.*, **41**, 289.
 Klinkenberg, L. J. 1941 *Drilling and production practice*, p. 200.
 Klose, W. 1931 *Ann. Phys., Lpz.*, **11**, 73.
 Knudsen, M. 1909 *Ann. Phys., Lpz.*, **28**, 75.
 Kohman, G. T. 1929 *J. Phys. Chem.* **33**, 229.
 Kozeny, J. 1927 *Wasserkr. u. Wasserwirt.* **22**, 67, 86.
 Krutter, H. & Day, J. R. 1941 *Oil Weekly*, **104** (4), 24.
 Loeb, L. B. 1934 *Kinetic theory of gases*, 2nd ed., p. 285.
 Millikan, R. 1923 *Phys. Rev.* **21**, 224.
 Rigden, P. 1947 *J. Soc. Chem. Ind., Lond.*, **66**, 130.
 von Smoluchowski, M. 1911 *Ann. Phys., Lpz.*, **35**, 983.

The theory of the transport phenomena in metals

By E. H. SONDHEIMER, Ph.D.

Fellow of Trinity College, Cambridge

(Communicated by A. H. Wilson, F.R.S.—Received 22 February 1950)

Exact expressions, valid for all temperatures, are obtained in the form of infinite determinants for the electrical conductivity, the thermal conductivity and the thermo-electric power of a degenerate gas of quasi-free electrons interacting with the ionic lattice of a metal. It is shown that the values of the electrical and thermal conductivities, in general, exceed the values given by the approximate interpolation formulae due to Bloch (1930), Wilson (1937) and others, and, in particular, that the Gruneisen-Bloch formula for the ideal electrical resistance is appreciably in error in the region close to the Debye temperature. It is further shown that the residual and ideal resistances of an impure metal are not strictly additive in the region where the two are of the same order of magnitude. The behaviour of the thermal conductivity is shown to agree qualitatively with the discussion based on Wilson's formula given by Makinson (1938); the numerical values of the thermal conductivity, however, are increased appreciably, particularly for an ideal metal at low temperatures. The thermo-electric power is also discussed, but no simple results can be given for the intermediate temperature range.

1. INTRODUCTION

The theoretical study of the transport phenomena in metals requires the solution of a complicated integral equation for the velocity distribution function of the conduction electrons. This equation has so far been solved only for the simplest model in which the electrons are assumed to be quasi-free; even for this case, however, no satisfactory and generally valid solution has yet been given, although much effort has been devoted towards obtaining solutions in a variety of special cases. At high temperatures, such that $(\Theta/T)^2$ can be neglected, where Θ is the Debye temperature, the integral equation reduces to an ordinary equation, and a solution is easily obtained (see, for example, Wilson 1936, p. 208); and the terms of higher order in $(\Theta/T)^2$ may be obtained by a method of successive approximations (Wilson 1937, § 2). The case of an impure metal at very low temperatures can also be treated by a method of successive approximations, which amounts to an expansion in descending powers of the residual resistance (Wilson 1937, § 3; Dube 1938); but the higher order terms become very complicated, and the method cannot be applied to the case of an ideally pure metal. The electrical conductivity of a pure metal can be obtained by a special method due to Bloch (1930), but this gives only the leading term at low temperatures, and it cannot be used at all to deal with the second-order effects such as the thermal conductivity and the thermo-electric power.

More general methods, applicable in principle to the whole temperature range, have been given by Kroll (1933 *a, b*) and more recently by Kohler (1948, 1949 *a*). Kroll transformed the integral equation into an infinite set of linear equations and obtained a solution by the use of infinite determinants; he confined himself to an evaluation of the leading terms only in the thermal conductivity and the thermo-electric power of an ideal metal at low temperatures.* Kohler transformed the

* It is shown in § 5 of the present paper that Kroll's original result for the thermal conductivity is, in fact, incorrect; see also Kroll (1938) and Umeda & Yamamoto (1949).

integral equation into a variational problem, and attempted to obtain a solution by expanding the distribution function as a power series and by using the variation principle to determine the coefficients. He evaluated explicitly only the approximations of lowest order, and found that they lead to expressions for the electrical and thermal conductivities which are identical with the interpolation formulae previously obtained by Wilson (1937) for an impure metal, assuming the general validity of Matthiessen's rule concerning the additivity of the residual and ideal resistances. (The expression for the ideal electrical resistance obtained by these methods is identical with the well-known Grüneisen-Bloch interpolation formula.) These formulae are, however, known to lead to incorrect results at intermediate temperatures (Dube 1938), and no explicit expressions for the transport magnitudes which are exact for all temperatures have so far been given.

Such expressions are provided in the present paper, which represents a synthesis and further development of Kroll's and Kohler's methods of solution. In § 2 the variational method as developed by Kohler (1949*a*) is recapitulated, and it is shown that it leads to equations which are identical with those obtained earlier by Kroll (1933*a*) by means of an arbitrary procedure.* In § 3 the calculations are completed without approximation, the transport magnitudes for a degenerate electron gas being finally obtained as the ratio of two infinite determinants (equations (38), (39) and (40)). The results may be evaluated numerically to any desired degree of accuracy by breaking off the determinants at a finite number of rows and columns, but no attempt is made to give a general discussion of the convergence of the method. Wilson's interpolation formulae are obtained on retaining only the lowest terms in the determinants which give a non-zero result, and it is shown in § 3.31 that the effect of the higher approximations is to increase the electrical and thermal conductivities above the values given by Wilson.

The electrical conductivity is discussed in detail in § 4, and numerical values are given. It is shown that the Grüneisen-Bloch formula leads to values of the ideal electrical conductivity which are appreciably too low at temperatures close to the Debye temperature, in agreement with the experimental facts. For an impure metal it is found that Matthiessen's rule breaks down in the important temperature region where the residual and ideal resistances are of the same order of magnitude. The deviations from the rule are small and positive, in qualitative agreement with experiment. A result obtained by Dube (1938), which seems to indicate that the deviations from Matthiessen's rule are negative, is discussed in § 4.22 and is shown to admit of no such interpretation.

The second-order phenomena are considered in § 5. It is shown that Wilson's formula for the thermal conductivity, discussed in detail by Makinson (1938), is qualitatively correct, and, in particular, the existence of a minimum at intermediate temperatures is confirmed. The numerical values of the thermal conductivity, however, are increased appreciably at intermediate and (for an ideal metal) at low tem-

* After completion of the present work, a paper by Umeda (1942) came to the author's notice in which the variational derivation of Kroll's equations is given independently of Kohler's work. Some particular results of the present paper have been anticipated by later Japanese work on Kroll's method (Umeda & Toya 1949; Umeda & Yamamoto 1949).

peratures. The paper concludes with a brief discussion of the thermo-electric power in § 5.2.

2. GENERAL THEORY

2.1. The integral equation

The discussion is restricted to monovalent metals and it is assumed that the electrons are quasi-free, so that the energy E is related to the wave-vector \mathbf{k} by $E = \hbar^2 |\mathbf{k}|^2 / (8\pi^2 m)$, m being the effective mass of an electron. In the presence of an electric field \mathcal{E} and a temperature gradient $\partial T / \partial x$ parallel to the x axis, the distribution function f of the conduction electrons is most conveniently written in the form

$$f = f_0 - k_1 c(\eta) \frac{\partial f_0}{\partial E}, \quad (1)$$

where $\eta = (E - \zeta) / kT$ (ζ being the Fermi energy level and k being Boltzmann's constant*), f_0 is the Fermi function $1 / (e^\eta + 1)$, and $c(\eta)$ is a function of η which has to be determined.

When scattering by both impurities and the lattice vibrations is taken into account, the integral equation for $c(\eta)$ is (Wilson 1937, equations (3) and (19))

$$4\Lambda \left(\frac{\hbar^2}{8\pi^2 m} \right)^{\frac{1}{2}} \left(\frac{\Theta}{T} \right)^3 E^{\frac{1}{2}} \left\{ \epsilon \mathcal{E} + T \frac{\partial}{\partial x} \left(\frac{\zeta}{T} \right) + \frac{E}{T} \frac{\partial T}{\partial x} \right\} = L(c), \quad (2)$$

where

$$L(c) = -\Lambda M \left(\frac{\Theta}{T} \right)^3 E^2 c(\eta) - \int_{-\Theta/T}^{\Theta/T} \left[E c(\eta) - c(\eta + z) \left\{ E + \frac{1}{2} k T z - D \left(\frac{T}{\Theta} \right)^2 z^2 \right\} \right] \times \frac{e^\eta + 1}{e^{\eta+z} + 1} \frac{z^2 dz}{|1 - e^{-z}|}, \quad (3)$$

$-\epsilon$ being the electronic charge, M being a constant depending on the number and scattering power of the impurities, and Λ and D being constants characteristic of the pure metal. The explicit form of $L(c)$ will not be needed for the general results to be proved in the present section.

2.2. The variation principle

Since the integral equation is linear, it is sufficient to solve it when the left-hand side is replaced by E^n , the corresponding solution being denoted by $c^{(n)}$. To obtain a solution use is made of a variational formulation due to Kohler (1948, 1949a). Kohler has shown that the correct solution $c^{(n)}$ is such as to make the integral†

$$(c^{(n)}, c^{(n)}) = \int_{-\infty}^{\infty} c^{(n)} L(c^{(n)}) \frac{\partial f_0}{\partial \eta} d\eta \quad (4)$$

* The use of k in two different senses, to denote the wave-vector \mathbf{k} with components k_1, k_2, k_3 , and to denote Boltzmann's constant k , should not cause any confusion.

† The relation between the present notation and that of Kohler (1949a) is most easily found by comparing equations (2) and (3) of the present paper with equations (1a), (1b) and (1c) of Kohler's paper. Kohler writes the curly bracket on the left-hand side of equation (2) in the form $(\epsilon \mathcal{E} + \partial \zeta / \partial x) + k\eta \partial T / \partial x$ and separates the integral equation accordingly; this procedure is in some respects more convenient than that used here, which is that used by Wilson (1937), but the final results are the same. Equations (4) and (5) of the present paper are essentially equivalent to equations (5a), (5b), (6a) and (6b) of Kohler's paper.

a maximum, subject to the subsidiary (normalizing) condition

$$\int_{-\infty}^{\infty} c^{(n)} L(c^{(n)}) \frac{\partial f_0}{\partial \eta} d\eta = \int_{-\infty}^{\infty} E^n c^{(n)} \frac{\partial f_0}{\partial \eta} d\eta. \quad (5)$$

He has also proved generally that, if ϕ_1 and ϕ_2 are any functions of η ,

$$(\phi_1, \phi_2) = (\phi_2, \phi_1), \quad (6)$$

and

$$(\phi_1, \phi_1) \geq 0. \quad (7)$$

2.21. Now expand $c^{(n)}$ as a power series in η ,

$$c^{(n)}(\eta) = \sum_{\mu=0}^{\infty} c_{\mu}^{(n)} \eta^{\mu}, \quad (8)$$

and use the variation principle to determine the coefficients. Equations (4) and (5) become

$$(c^{(n)}, c^{(n)}) = \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} d_{\mu\nu} c_{\mu}^{(n)} c_{\nu}^{(n)}. \quad (9)$$

$$\text{and} \quad \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} d_{\mu\nu} c_{\mu}^{(n)} c_{\nu}^{(n)} = \sum_{\mu=0}^{\infty} \alpha_{\mu}^{(n)} c_{\mu}^{(n)} \quad (10)$$

$$\text{where}^* \quad d_{\mu\nu} = (\eta^{\mu}, \eta^{\nu}) = \int_{-\infty}^{\infty} \eta^{\mu} L(\eta^{\nu}) \frac{\partial f_0}{\partial \eta} d\eta = d_{\nu\mu}, \quad (11)$$

$$\text{and} \quad \alpha_{\mu}^{(n)} = \int_{-\infty}^{\infty} E^n \eta^{\mu} \frac{\partial f_0}{\partial \eta} d\eta. \quad (12)$$

By differentiating with respect to $c_{\mu}^{(n)}$ (say), it is found that the maximum value of (9), subject to the condition (10), is obtained if the coefficients $c_{\mu}^{(n)}$ satisfy the infinite set of equations

$$\sum_{\nu=0}^{\infty} d_{\mu\nu} c_{\nu}^{(n)} = \alpha_{\mu}^{(n)} \quad (\mu = 0, 1, 2, \dots). \quad (13)$$

Equation (13) could have been obtained formally from equation (2) by multiplying by $\eta^{\mu} \partial f_0 / \partial \eta$, where $\mu = 0, 1, 2, \dots$, and integrating with respect to η . This is, in fact, the procedure adopted (without special justification) by Kroll (1933 *a*; see also Wilson 1936, § 6.5).

2.3. The current densities

If \mathbf{v} is the velocity of an electron ($\hbar \mathbf{v} = 2\pi \text{ grad}_{\mathbf{k}} E$), the electric current is given by

$$\begin{aligned} J &= -\frac{e}{4\pi^3} \iiint v_x f dk_1 dk_2 dk_3 = \frac{16\pi^2}{3\hbar^4} (2m)^{\frac{3}{2}} e \int_{-\infty}^{\infty} E^{\frac{1}{2}} c(\eta) \frac{\partial f_0}{\partial \eta} d\eta \\ &= \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} \left\{ e^2 \mathcal{E} + eT \frac{\partial}{\partial x} \left(\frac{\xi}{T} \right) \right\} + \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} \frac{e}{T} \frac{\partial T}{\partial x}, \end{aligned} \quad (14)$$

and the heat current is given by

$$\begin{aligned} w &= \frac{1}{4\pi^3} \iiint v_x E f dk_1 dk_2 dk_3 \\ &= \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} \left\{ -e\mathcal{E} - T \frac{\partial}{\partial x} \left(\frac{\xi}{T} \right) \right\} - \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} \frac{1}{T} \frac{\partial T}{\partial x}, \end{aligned} \quad (15)$$

* Apart from a multiplicative constant, the quantity $d_{\mu\nu}$ defined here is the same as that used by Kohler (1949*a*, equation (8*c*)). Also $\alpha_{\mu}^{(1)}$ and $\alpha_{\mu}^{(2)}$ correspond to Kohler's α_{μ} and β_{μ} , but are not identical with them for the reasons given in the footnote on p. 77.

where, using equations (8) and (12),

$$\mathcal{K}_{m,n} = \frac{8\Lambda}{3\pi\hbar} \left(\frac{\Theta}{T}\right)^3 \int_{-\infty}^{\infty} E^m c^{(n)}(\eta) \frac{\partial f_0}{\partial \eta} d\eta = \frac{8\Lambda}{3\pi\hbar} \left(\frac{\Theta}{T}\right)^3 \sum_{\mu=0}^{\infty} \alpha_{\mu}^{(m)} c_{\mu}^{(n)} = \mathcal{K}_{n,m}. \quad (16)$$

(Note that the notation for the \mathcal{K} 's differs slightly from that used by Wilson (1937).)

2.31. Solving the infinite set of equations (13) for the $c_{\mu}^{(n)}$ and substituting the result in (16), it is easily shown that

$$\mathcal{K}_{m,n} = -\frac{8\Lambda}{3\pi\hbar} \left(\frac{\Theta}{T}\right)^3 \frac{\mathcal{D}_{m,n}}{\mathcal{D}}, \quad (17)$$

where \mathcal{D} is the determinant of the $d_{\mu\nu}$'s,

$$\mathcal{D} = |d_{\mu\nu}|, \quad (18)$$

and where

$$\mathcal{D}_{m,n} = \begin{vmatrix} 0 & \alpha_0^{(m)} & \alpha_1^{(m)} & \alpha_2^{(m)} & \dots \\ \alpha_0^{(n)} & d_{00} & d_{01} & d_{02} & \dots \\ \alpha_1^{(n)} & d_{10} & d_{11} & d_{12} & \dots \\ \alpha_2^{(n)} & d_{20} & d_{21} & d_{22} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \quad (19)$$

All the transport effects of interest are readily expressed in terms of the $\mathcal{K}_{m,n}$, and hence in terms of the quantities $d_{\mu\nu}$ and $\alpha_{\mu}^{(n)}$.

The electrical conductivity is given by

$$\sigma = e^2 \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} = -\frac{8e^2\Lambda}{3\pi\hbar} \left(\frac{\Theta}{T}\right)^3 \frac{\mathcal{D}_{\frac{1}{2}, \frac{1}{2}}}{\mathcal{D}}. \quad (20)$$

The thermal conductivity is

$$\kappa = -\frac{\mathcal{K}_{\frac{3}{2}, \frac{1}{2}} \mathcal{K}_{\frac{1}{2}, \frac{1}{2}} - \mathcal{K}_{\frac{1}{2}, \frac{3}{2}}^2}{\mathcal{K}_{\frac{1}{2}, \frac{1}{2}} T} = \frac{8\Lambda}{3\pi\hbar\Theta} \left(\frac{\Theta}{T}\right)^4 \left(-\frac{\mathcal{D}_{\frac{3}{2}, \frac{1}{2}}}{\mathcal{D}} + \frac{\mathcal{D}_{\frac{1}{2}, \frac{1}{2}}^2}{\mathcal{D}_{\frac{1}{2}, \frac{3}{2}}} \right). \quad (21a)$$

As shown by Kohler, Sylvester's theorem (see, for example, Kowalewski 1909) may be used to write this in the simpler form

$$\kappa = -\frac{8\Lambda}{3\pi\hbar\Theta} \left(\frac{\Theta}{T}\right)^4 \frac{\mathcal{D}_{\frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}}}{\mathcal{D}_{\frac{1}{2}, \frac{1}{2}}}, \quad (21b)$$

where

$$\mathcal{D}_{\frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}} = \begin{vmatrix} 0 & 0 & \alpha_0^{(\frac{3}{2})} & \alpha_1^{(\frac{3}{2})} & \alpha_2^{(\frac{3}{2})} & \dots \\ 0 & 0 & \alpha_0^{(\frac{1}{2})} & \alpha_1^{(\frac{1}{2})} & \alpha_2^{(\frac{1}{2})} & \dots \\ \alpha_0^{(\frac{3}{2})} & \alpha_0^{(\frac{1}{2})} & d_{00} & d_{01} & d_{02} & \dots \\ \alpha_1^{(\frac{3}{2})} & \alpha_1^{(\frac{1}{2})} & d_{10} & d_{11} & d_{12} & \dots \\ \alpha_2^{(\frac{3}{2})} & \alpha_2^{(\frac{1}{2})} & d_{20} & d_{21} & d_{22} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \quad (22)$$

Finally, the absolute thermo-electric power per degree is $-\mathcal{E}/e$, where

$$\mathcal{E} = \frac{\mathcal{K}_{\frac{3}{2}, \frac{1}{2}} - \zeta \mathcal{K}_{\frac{1}{2}, \frac{1}{2}}}{\mathcal{K}_{\frac{1}{2}, \frac{1}{2}} T} = \frac{\mathcal{D}_{\frac{3}{2}, \frac{1}{2}} - \zeta \mathcal{D}_{\frac{1}{2}, \frac{1}{2}}}{\mathcal{D}_{\frac{1}{2}, \frac{1}{2}} T}. \quad (23)$$

3. EXPLICIT EVALUATION OF THE TRANSPORT MAGNITUDES FOR A DEGENERATE ELECTRON GAS

3.1. The quantities $\alpha_\mu^{(n)}$ and $d_{\mu\nu}$, defined by equations (12) and (11) (L being the integral operator (3)), may be evaluated by expanding in powers of η and using the usual methods for integrals involving the Fermi distribution function. The calculations are straightforward but (in the case of the $d_{\mu\nu}$) somewhat lengthy; details are given in appendix 1, and only the final results will be quoted here.

It is found that

$$\alpha_\mu^{(n)} = -\zeta^n a_\mu - \zeta^n \sum_{r=2,4,6,\dots}^{\infty} \frac{n(n-1)(n-2)\dots(n-r+1)}{r!} a_{\mu+r} \gamma^r \quad (24a)$$

if μ is even (including $\mu = 0$), and that

$$\alpha_\mu^{(n)} = -\zeta^n \sum_{r=1,3,5,\dots}^{\infty} \frac{n(n-1)(n-2)\dots(n-r+1)}{r!} a_{\mu+r} \gamma^r \quad (24b)$$

if μ is odd, where $\gamma = kT/\zeta$ is the degeneracy parameter, and where

$$\left. \begin{aligned} a_0 &= 1, \\ a_{2s} &= 2(2s)! \sum_{l=1}^{\infty} \frac{(-1)^{l+1}}{l^{2s}} = (2^{2s} - 2) B_s \pi^{2s} \quad (s \geq 1), \end{aligned} \right\} \quad (25)$$

B_s being the s th Bernoulli number.

$$\text{Further,} \quad d_{\mu\nu} = \Lambda M \left(\frac{\Theta}{T} \right)^3 \zeta^2 (a_{\mu+\nu} + \gamma^2 a_{\mu+\nu+2}) + 2D \left(\frac{T}{\Theta} \right)^2 F_{\mu\nu} \quad (26a)$$

if $\mu + \nu$ is even, and

$$d_{\mu\nu} = 2\gamma \Lambda M \left(\frac{\Theta}{T} \right)^3 \zeta^2 a_{\mu+\nu+1} + 2\gamma D \left(\frac{T}{\Theta} \right)^2 F_{\mu\nu} \quad (26b)$$

if $\mu + \nu$ is odd, where

$$\begin{aligned} F_{\mu\nu} &= a_{\mu+\nu} \mathcal{J}_5 + \sum_{k=3,5,7,\dots}^{\mu-1,\mu} \binom{\mu}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} \left\{ \mathcal{J}_{k+4} - \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \mathcal{J}_{k+2} \right\} \\ &+ \sum_{k=3,5,7,\dots}^{\nu-1,\nu} \binom{\nu}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} \left\{ \mathcal{J}_{k+4} - \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \mathcal{J}_{k+2} \right\} \\ &+ \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \sum_{k=3,5,7,\dots}^{\mu+\nu+1} \binom{\mu+\nu+1}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu+1} \mathcal{J}_{k+2} \\ &+ (-1)^\mu \frac{\mu! \nu!}{(\mu+\nu+1)!} \left\{ \mathcal{J}_{\mu+\nu+5} - \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \mathcal{J}_{\mu+\nu+3} \right\} \end{aligned} \quad (27a)$$

if $\mu + \nu$ is even, and where

$$\begin{aligned} F_{\mu\nu} &= \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left[\left\{ \frac{1}{4}(\mu+\nu) + \frac{1}{2}\mu\nu \right\} a_{\mu+\nu-1} \mathcal{J}_5 \right. \\ &- \sum_{k=4,6,8,\dots}^{\mu-1,\mu} \left\{ \left(\frac{\mu+1}{k+1} \right) - \frac{1}{2} \binom{\mu}{k} \right\} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} \mathcal{J}_{k+3} \\ &- \sum_{k=4,6,8,\dots}^{\nu-1,\nu} \left\{ \left(\frac{\nu+1}{k+1} \right) - \frac{1}{2} \binom{\nu}{k} \right\} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} \mathcal{J}_{k+3} \\ &\left. + \sum_{k=4,6,8,\dots}^{\mu+\nu+1} \binom{\mu+\nu+2}{k+1} \frac{a_{\mu+\nu-k+1}}{\mu+\nu+2} \mathcal{J}_{k+3} + \frac{1}{2} (-1)^\mu (\mu-\nu) \frac{\mu! \nu!}{(\mu+\nu+2)!} \mathcal{J}_{\mu+\nu+4} \right] \end{aligned} \quad (27b)$$

if $\mu + \nu$ is odd. In these expressions the symbol $\sum_{k=3,5,7,\dots}^{\mu-1,\mu}$, for example, means that the upper limit of the summation is to be taken as $\mu - 1$ or μ according as μ is even or odd, and, moreover, that the summation is to be omitted entirely if $\mu < 3$. It is evident by inspection that $F_{\mu\nu} = F_{\nu\mu}$; as particular cases, we have

$$F_{0\nu} = F_{\nu 0} = \sum_{k=1,3,5,\dots}^{\nu+1} \binom{\nu+1}{k} \frac{\alpha_{\nu-k+1}}{\nu+1} \mathcal{J}_{k+4} \quad (28a)$$

$$\text{if } \nu \text{ is even, and } F_{0\nu} = F_{\nu 0} = \frac{1}{2} \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \sum_{k=2,4,6,\dots}^{\nu+1} \binom{\nu+1}{k} \frac{\alpha_{\nu-k+1}}{\nu+1} \mathcal{J}_{k+3} \quad (28b)$$

if ν is odd. Finally, in the above expressions,

$$\mathcal{J}_n = \int_0^{\Theta/T} \frac{z^n dz}{(e^z - 1)(1 - e^{-z})}. \quad (29)$$

It should be particularly noted that the calculations have been carried out retaining *all* powers of γ . The expressions for $\alpha_\mu^{(n)}$ are infinite series in ascending powers of γ^2 , while the $d_{\mu\nu}$ involve powers of γ up to and including the second only.

3.2. The transport magnitudes can now be evaluated explicitly by substituting the expressions (24) and (26) in equations (20), (21) and (23). For a degenerate electron gas it is sufficient to retain only the lowest terms in an expansion in ascending powers of γ^2 . This enables us to simplify the resulting expressions considerably; it should be noted, however, that this approximation does *not*, in general, allow us to break off the infinite determinants at a finite number of rows and columns.

It follows from (18) and (19), for example, that the leading term in the determinant \mathcal{D} is independent of γ , and is

$$\mathcal{D} = \begin{vmatrix} d'_{00} & 0 & d'_{02} & 0 & d'_{04} & \dots \\ 0 & d'_{11} & 0 & d'_{13} & 0 & \dots \\ d'_{20} & 0 & d'_{22} & 0 & d'_{24} & \dots \\ 0 & d'_{31} & 0 & d'_{33} & 0 & \dots \\ d'_{40} & 0 & d'_{42} & 0 & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix},$$

where $d'_{\mu\nu}$ is the coefficient of the lowest power of γ in $d_{\mu\nu}$. Rearranging rows and columns, it is found that

$$\mathcal{D} = \begin{vmatrix} d'_{11} & d'_{13} & d'_{15} & \dots & 0 & 0 & 0 & \dots \\ d'_{31} & d'_{33} & d'_{35} & \dots & 0 & 0 & 0 & \dots \\ d'_{51} & d'_{53} & d'_{55} & \dots & 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \dots & d'_{00} & d'_{02} & d'_{04} & \dots \\ 0 & 0 & 0 & \dots & d'_{20} & d'_{22} & d'_{24} & \dots \\ 0 & 0 & 0 & \dots & d'_{40} & d'_{42} & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} = \begin{vmatrix} d'_{11} & d'_{13} & d'_{15} & \dots \\ d'_{31} & d'_{33} & d'_{35} & \dots \\ d'_{51} & d'_{53} & d'_{55} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} \times \begin{vmatrix} d'_{00} & d'_{02} & d'_{04} & \dots \\ d'_{20} & d'_{22} & d'_{24} & \dots \\ d'_{40} & d'_{42} & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \quad (30)$$

Similarly,

$$\mathcal{D}_{\frac{1}{2}, \frac{1}{2}} = \zeta^3 \begin{vmatrix} 0 & a_0 & 0 & a_2 & 0 & \dots \\ a_0 & d'_{00} & 0 & d'_{02} & 0 & \dots \\ 0 & 0 & d'_{11} & 0 & d'_{13} & \dots \\ a_2 & d'_{20} & 0 & d'_{22} & 0 & \dots \\ 0 & 0 & d'_{31} & 0 & d'_{33} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} = \zeta^3 \begin{vmatrix} d'_{11} & d'_{13} & d'_{15} & \dots \\ d'_{31} & d'_{33} & d'_{35} & \dots \\ d'_{51} & d'_{53} & d'_{55} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} \times \begin{vmatrix} 0 & a_0 & a_2 & a_4 & \dots \\ a_0 & d'_{00} & d'_{02} & d'_{04} & \dots \\ a_2 & d'_{20} & d'_{22} & d'_{24} & \dots \\ a_4 & d'_{40} & d'_{42} & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \quad (31)$$

and, after somewhat more complicated manipulations,

$$\mathcal{D}_{\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}} = \gamma^2 \zeta^8 \begin{vmatrix} 0 & a_2 & a_4 & a_6 & \dots \\ a_2 & d'_{11} & d'_{13} & d'_{15} & \dots \\ a_4 & d'_{31} & d'_{33} & d'_{35} & \dots \\ a_6 & d'_{51} & d'_{53} & d'_{55} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix} \times \begin{vmatrix} 0 & a_0 & a_2 & a_4 & \dots \\ a_0 & d'_{00} & d'_{02} & d'_{04} & \dots \\ a_2 & d'_{20} & d'_{22} & d'_{24} & \dots \\ a_4 & d'_{40} & d'_{42} & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \quad (32)$$

and

$$\mathcal{D}_{\frac{1}{2}, \frac{1}{2}} - \zeta \mathcal{D}_{\frac{1}{2}, \frac{1}{2}} = \gamma^2 \zeta^4 \begin{vmatrix} 0 & a_0 & \frac{3}{2}a_2 & a_2 & \frac{3}{2}a_4 & a_4 & \dots \\ \frac{3}{2}a_2 & d'_{00} & d'_{01} & d'_{02} & d'_{03} & d'_{04} & \dots \\ a_2 & 0 & d'_{11} & 0 & d'_{13} & 0 & \dots \\ \frac{3}{2}a_4 & d'_{20} & d'_{21} & d'_{22} & d'_{23} & d'_{24} & \dots \\ a_4 & 0 & d'_{31} & 0 & d'_{33} & 0 & \dots \\ \frac{3}{2}a_6 & d'_{40} & d'_{41} & d'_{42} & d'_{43} & d'_{44} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \quad (33)$$

only the lowest power of γ^2 being retained in each case.

3.3. The final results are most conveniently expressed in terms of the approximate interpolation formulae for the electrical and thermal conductivities of an impure metal given by Wilson (1937, p. 378). From (26) it is readily shown that

$$d'_{\mu\nu} = \frac{8e^2\zeta^3\Lambda}{3\pi h} \left(\frac{\Theta}{T}\right)^3 \left(\frac{a_{\mu+\nu}}{\sigma_r} + \frac{F_{\mu\nu}}{\mathcal{J}_5\sigma_i}\right) \quad (34a)$$

$$\text{if } \mu + \nu \text{ is even, and} \quad d'_{\mu\nu} = \frac{8e^2\zeta^3\Lambda}{3\pi h} \left(\frac{\Theta}{T}\right)^3 \left(\frac{2a_{\mu+\nu+1}}{\sigma_r} + \frac{F_{\mu\nu}}{\mathcal{J}_5\sigma_i}\right) \quad (34b)$$

if $\mu + \nu$ is odd, where $1/\sigma_r = \rho_r = 3\pi h M / (8e^2\zeta)$ is the residual electrical resistance, and where

$$\frac{1}{\sigma_i} = \rho_i = \frac{3\pi h D}{4e^2\zeta^3\Lambda} \left(\frac{T}{\Theta}\right)^5 \mathcal{J}_5 \quad (35)$$

is the Grüneisen-Bloch interpolation formula for the ideal electrical resistance. Alternatively, for $\mu + \nu$ even, $d'_{\mu\nu}$ may be written as

$$d'_{\mu\nu} = \frac{8\pi k^2\zeta^3\Lambda\Theta}{9h} \left(\frac{\Theta}{T}\right)^2 \left(\frac{a_{\mu+\nu}}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{\mu\nu}}{F_{11}\kappa_i}\right) \quad (36)$$

(F_{11} being obtained from (27a)), where $\kappa_r = \frac{1}{3}(\pi k/\epsilon)^2 \sigma_r T$ is the residual thermal conductivity (note that the present notation differs slightly from Wilson's), and where

$$\frac{1}{\kappa_i} = \frac{27h}{8\pi^3 k^2 \zeta^2 \Lambda \Theta} \left(\frac{T}{\Theta} \right)^2 \left\{ \mathcal{J}_5 + \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 \left(\frac{2\pi^2}{3} \mathcal{J}_5 - \frac{1}{3} \mathcal{J}_7 \right) \right\} \quad (37)$$

is Wilson's interpolation formula for the ideal part of the thermal resistance.*

Combining equations (20), (30), (31) and (34a) and removing a common factor from numerator and denominator, the final expression for the electrical conductivity is obtained in the form

$$\sigma = - \frac{\begin{vmatrix} 0 & a_0 & a_2 & a_4 & \dots \\ a_0 & \frac{1}{\sigma_r} + \frac{1}{\sigma_i} & \frac{a_2}{\sigma_r} + \frac{F_{02}}{\mathcal{J}_5 \sigma_i} & \frac{a_4}{\sigma_r} + \frac{F_{04}}{\mathcal{J}_5 \sigma_i} & \dots \\ a_2 & \frac{a_2}{\sigma_r} + \frac{F_{20}}{\mathcal{J}_5 \sigma_i} & \frac{a_4}{\sigma_r} + \frac{F_{22}}{\mathcal{J}_5 \sigma_i} & \frac{a_6}{\sigma_r} + \frac{F_{24}}{\mathcal{J}_5 \sigma_i} & \dots \\ a_4 & \frac{a_4}{\sigma_r} + \frac{F_{40}}{\mathcal{J}_5 \sigma_i} & \frac{a_6}{\sigma_r} + \frac{F_{42}}{\mathcal{J}_5 \sigma_i} & \frac{a_8}{\sigma_r} + \frac{F_{44}}{\mathcal{J}_5 \sigma_i} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}}{\begin{vmatrix} \frac{1}{\sigma_r} + \frac{1}{\sigma_i} & \frac{a_2}{\sigma_r} + \frac{F_{02}}{\mathcal{J}_5 \sigma_i} & \frac{a_4}{\sigma_r} + \frac{F_{04}}{\mathcal{J}_5 \sigma_i} & \dots \\ \frac{a_2}{\sigma_r} + \frac{F_{20}}{\mathcal{J}_5 \sigma_i} & \frac{a_4}{\sigma_r} + \frac{F_{22}}{\mathcal{J}_5 \sigma_i} & \frac{a_6}{\sigma_r} + \frac{F_{24}}{\mathcal{J}_5 \sigma_i} & \dots \\ \frac{a_4}{\sigma_r} + \frac{F_{40}}{\mathcal{J}_5 \sigma_i} & \frac{a_6}{\sigma_r} + \frac{F_{42}}{\mathcal{J}_5 \sigma_i} & \frac{a_8}{\sigma_r} + \frac{F_{44}}{\mathcal{J}_5 \sigma_i} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}}, \quad (38)$$

the a_μ 's being given by (25) and the $F_{\mu\nu}$'s by (27a).

Similarly, combining (21b), (31), (32) and (36), the thermal conductivity is given as

$$\kappa = - \frac{3}{\pi^2} \frac{\begin{vmatrix} 0 & a_2 & a_4 & a_6 & \dots \\ a_2 & \frac{1}{3}\pi^2 \left(\frac{1}{\kappa_r} + \frac{1}{\kappa_i} \right) & \frac{a_4}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{13}}{F_{11}\kappa_i} & \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{15}}{F_{11}\kappa_i} & \dots \\ a_4 & \frac{a_4}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{31}}{F_{11}\kappa_i} & \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{33}}{F_{11}\kappa_i} & \frac{a_8}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{35}}{F_{11}\kappa_i} & \dots \\ a_6 & \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{51}}{F_{11}\kappa_i} & \frac{a_8}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{53}}{F_{11}\kappa_i} & \frac{a_{10}}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{55}}{F_{11}\kappa_i} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}}{\begin{vmatrix} \frac{1}{3}\pi^2 \left(\frac{1}{\kappa_r} + \frac{1}{\kappa_i} \right) & \frac{a_4}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{13}}{F_{11}\kappa_i} & \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{15}}{F_{11}\kappa_i} & \dots \\ \frac{a_4}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{31}}{F_{11}\kappa_i} & \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{33}}{F_{11}\kappa_i} & \frac{a_8}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{35}}{F_{11}\kappa_i} & \dots \\ \frac{a_6}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{51}}{F_{11}\kappa_i} & \frac{a_8}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{53}}{F_{11}\kappa_i} & \frac{a_{10}}{\kappa_r} + \frac{1}{3}\pi^2 \frac{F_{55}}{F_{11}\kappa_i} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}}, \quad (39)$$

* This formula is introduced here for mathematical convenience; we may note, however (see § 5 below), that, unlike equation (35), it *cannot* be used as an interpolation formula for the ideal resistance in the absence of impurities.

and, combining (23), (31), (33) and (34), the thermo-electric power is given as

$$\mathfrak{S} = \frac{k^2 T}{\zeta} \begin{vmatrix} 0 & a_0 & \frac{3}{2}a_2 & a_2 & \frac{3}{2}a_4 & \dots \\ \frac{3}{2}a_2 & \frac{1}{\sigma_r} + \frac{1}{\sigma_i} & \frac{2a_2}{\sigma_r} + \frac{F_{01}}{\mathcal{J}_5 \sigma_i} & \frac{a_2}{\sigma_r} + \frac{F_{02}}{\mathcal{J}_5 \sigma_i} & \frac{2a_4}{\sigma_r} + \frac{F_{03}}{\mathcal{J}_5 \sigma_i} & \dots \\ a_2 & 0 & \frac{a_2}{\sigma_r} + \frac{F_{11}}{\mathcal{J}_5 \sigma_i} & 0 & \frac{a_4}{\sigma_r} + \frac{F_{13}}{\mathcal{J}_5 \sigma_i} & \dots \\ \frac{3}{2}a_4 & \frac{a_2}{\sigma_r} + \frac{F_{20}}{\mathcal{J}_5 \sigma_i} & \frac{2a_4}{\sigma_r} + \frac{F_{21}}{\mathcal{J}_5 \sigma_i} & \frac{a_4}{\sigma_r} + \frac{F_{22}}{\mathcal{J}_5 \sigma_i} & \frac{2a_6}{\sigma_r} + \frac{F_{23}}{\mathcal{J}_5 \sigma_i} & \dots \\ a_4 & 0 & \frac{a_4}{\sigma_r} + \frac{F_{31}}{\mathcal{J}_5 \sigma_i} & 0 & \frac{a_6}{\sigma_r} + \frac{F_{33}}{\mathcal{J}_5 \sigma_i} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}. \quad (40)$$

3.31. An alternative form for σ and κ , which is particularly convenient for numerical evaluation, may be obtained as follows. Let $S^{(2n)}$ be the determinant obtained by breaking off the denominator of (38) at the $(n+1)$ th row and column, the element in the bottom right-hand corner being $\frac{a_{4n}}{\sigma_r} + \frac{F_{2n,2n}}{\mathcal{J}_5 \sigma_i}$. Also let $S_{a,a}^{(2n)}$ be the determinant of $n+2$ rows and columns obtained by bordering $S^{(2n)}$ with 0, a_0 , a_2 , ..., a_{2n} , and let $S_a^{(2n)}$ be the determinant obtained from $S^{(2n)}$ on replacing the last column by a_0 , a_2 , ..., a_{2n} . Then Sylvester's theorem gives

$$\begin{vmatrix} S^{(2n)} & S_a^{(2n)} \\ S_a^{(2n)} & S_{a,a}^{(2n-2)} \end{vmatrix} = S^{(2n-2)} S_{a,a}^{(2n)},$$

or, rearranging,
$$-\frac{S_{a,a}^{(2n)}}{S^{(2n)}} + \frac{S_{a,a}^{(2n-2)}}{S^{(2n-2)}} = \frac{(S_a^{(2n)})^2}{S^{(2n-2)} S^{(2n)}}. \quad (41)$$

Also
$$-\frac{S_{a,a}^{(0)}}{S^{(0)}} = \sigma^{(0)}, \quad (42)$$

where
$$\frac{1}{\sigma^{(0)}} = \frac{1}{\sigma_r} + \frac{1}{\sigma_i} \quad (43)$$

is the expression for the electrical resistance obtained by assuming the general validity of Matthiessen's rule, σ , being given by (35).

In (41) put $n = 1, 2, 3, \dots$, and add each of the resulting expressions to (42). This gives

$$\sigma = \sigma^{(0)} + \sum_{n=1}^{\infty} \frac{(S_a^{(2n)})^2}{S^{(2n-2)} S^{(2n)}}. \quad (44)$$

Similarly, for the thermal conductivity,

$$\kappa = \kappa^{(0)} + \frac{3}{\pi^2} \sum_{n=1}^{\infty} \frac{(K_a^{(2n)})^2}{K^{(2n-2)} K^{(2n)}}, \quad (45)$$

where

$$\frac{1}{\kappa^{(0)}} = \frac{1}{\kappa_r} + \frac{1}{\kappa_i}, \quad (46)$$

κ_i being given by (37), and where $K^{(2n)}$ is the determinant obtained by breaking off the denominator of (39) at the $(n+1)$ th row and column, while $K_a^{(2n)}$ is obtained from $K^{(2n)}$ on replacing the last column by $a_2, a_4, \dots, a_{2n+2}$.

In equations (44) and (45) the correction terms to Wilson's interpolation formulae are exhibited explicitly. It may be shown that the determinants $S^{(2n)}$ and $K^{(2n)}$ are positive (this is essentially due to the fact that the $d_{\mu\nu}$ defined by (11) are the coefficients of a positive definite quadratic form), and the series (44) and (45) are therefore monotonically increasing sequences.

4. THE ELECTRICAL CONDUCTIVITY

4.1. The expressions (38) and (44) for the electrical conductivity are valid for all temperatures (provided that it is permissible to treat the electron gas as completely degenerate), and for ideally pure as well as for impure metals. Previous solutions have all been confined to the regions of temperatures high and low compared with the Debye temperature, for which the complicated general expressions reduce to comparatively simple ones. At high temperatures, Wilson (1937, § 2) has shown that the conductivity is given by (43), correct to terms of order $(\Theta/T)^2$ and $M\Theta/T$.* Further, Bloch (1930) has shown that the ideal part (35) of (43) represents the leading term in the expression for the conductivity of an ideal metal at very low temperatures, and Wilson (1937, § 3) has shown that (43) also gives the conductivity of an impure metal at low temperatures, correct to terms of order σ_r/σ_i . (These special results can, of course, be shown to follow directly from (38) and (44), the proofs will not be given here.) At sufficiently high and low temperatures, therefore, $\sigma = \sigma^{(0)}$, Matthiessen's rule is valid, and the correction terms represented by the infinite sum in equation (44) are negligible. At intermediate temperatures, however, such that T is comparable with Θ , or σ_r with σ_i , the correction terms become appreciable and increase the electrical conductivity above the value given by (43). For an ideal metal, each correction term is proportional to $(\Theta/T)^5$ at high temperatures and to $(\Theta/T)^3$ at low temperatures ($\sigma^{(0)}$ being proportional to Θ/T and $(\Theta/T)^5$ respectively); while, for an impure metal, each correction term is proportional to $(T/\Theta)^{10}/M^3$ at low temperatures, $\sigma^{(0)}$ being proportional to $1/M$. It is, in general, no longer possible to separate the resistivity into a part depending upon ρ_r only and a part depending

* In discussing the high-temperature limit Wilson actually confined himself to the case of an ideally pure metal ($M = 0$), but his method is easily extended to include the impurity term.

upon ρ_i only; no simple formula can be given for the deviations from Matthiessen's rule, but they are evaluated numerically below for some typical cases.

4.2. The magnitude of the correction terms has to be estimated by numerical methods. For this purpose it is convenient to rewrite equation (44) in terms of the notation introduced by Makinson (1938). The ideal electrical resistance at high temperatures, such that $(\Theta/T)^2$ can be neglected compared with unity, is (Wilson 1936, p. 208)

$$1/\sigma_i = AT/\Theta, \quad (47)$$

where

$$A = 3\pi\hbar D/(16e^2\zeta^3\Lambda). \quad (48)$$

With this notation, (44) gives

$$4A\sigma = \frac{4A}{\rho} = \frac{1}{\left(\frac{T}{\Theta}\right)^5 \mathcal{J}_5 + \frac{\rho_r}{4A}} + \sum_{n=1}^{\infty} \frac{(\xi_a^{(2n)})^2}{\xi^{(2n-2)} \xi^{(2n)}}, \quad (49)$$

where

$$\xi^{(2n)} = \begin{vmatrix} \left(\frac{T}{\Theta}\right)^5 F_{00} + \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{02} + a_2 \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{0,2n} + a_{2n} \frac{\rho_r}{4A} \\ \left(\frac{T}{\Theta}\right)^5 F_{20} + a_2 \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{22} + a_4 \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{2,2n} + a_{2n+2} \frac{\rho_r}{4A} \\ \dots & \dots & \dots & \dots \\ \left(\frac{T}{\Theta}\right)^5 F_{2n,0} + a_{2n} \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{2n,2} + a_{2n+2} \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{2n,2n} + a_{4n} \frac{\rho_r}{4A} \end{vmatrix}, \quad (50)$$

and where $\xi_a^{(2n)}$ is obtained from $\xi^{(2n)}$ on replacing the last column by a_0, a_2, \dots, a_{2n} . The α_μ 's are given by (25) and the $F_{\mu\nu}$'s by (27a) and (28a). The parameter $\rho_r/4A$ is a measure of the amount of impurity present, and it is seen that the quantity $4A\sigma$ depends upon T/Θ , D/ζ and $\rho_r/4A$ only. For monovalent metals, moreover, which alone are considered in the present paper, $D/\zeta = 2^{-\frac{1}{2}}$ (Makinson 1938).

4.21. Equation (49) may be used to compute successive approximations to σ by retaining successive terms of the infinite series; the value obtained by retaining N terms of the series will be denoted by $\sigma^{(2N)} = 1/\rho^{(2N)}$. In the approximation of zero order, Matthiessen's rule is valid and the resistance is given by (43).* If $\rho_i^{(2N)}$ denotes the resistance of the ideal metal in the N th approximation, the deviation $\Delta^{(2N)}$ from Matthiessen's rule is

$$\Delta^{(2N)} = \rho^{(2N)} - (\rho_i^{(2N)} + \rho_r); \quad (51)$$

in general it will depend both upon the temperature and upon the amount of impurity present.

The first three approximations have been evaluated numerically, table 1 giving values of $\rho_i^{(2N)}/4A$ for the pure metal, and of $\rho^{(2N)}/4A$ and $\Delta^{(2N)}/4A$ for two values of the impurity parameter $\rho_r/4A$. Explicit expressions for the α_μ 's and the $F_{\mu\nu}$'s required in these calculations, and a table of values of the integrals \mathcal{J}_n defined by (29), are given in appendix 2.

* Kohler (1949a) attempted to evaluate the first correction term, and claimed to show that it is zero for a completely degenerate electron gas. This (incorrect) result is due to an error in Kohler's evaluation of the quantity d_{22} (equation (18) of Kohler's paper).

TABLE 1. THE ELECTRICAL RESISTANCE OF MONOVALENT METALS

$T/^\circ$	$100\rho_r/4A = 0$				$100\rho_r/4A = 1$				$100\rho_r/4A = 3$			
	$100 \frac{\rho_1^{(0)}}{4A}$	$100 \frac{\rho_1^{(2)}}{4A}$	$100 \frac{\rho_1^{(4)}}{4A}$		$100 \frac{\rho^{(0)}}{4A}$	$100 \frac{\rho^{(2)}}{4A}$	$100 \frac{\rho^{(4)}}{4A}$		$100 \frac{\rho^{(0)}}{4A}$	$100 \frac{\rho^{(2)}}{4A}$	$100 \frac{\rho^{(4)}}{4A}$	$100 \frac{\Delta^{(4)}}{4A}$
0.01	1.2443×10^{-6}	1.2429×10^{-6}	1.2429×10^{-6}		—	—	—	—	—	—	—	—
0.025	1.2152×10^{-4}	1.2068×10^{-4}	1.2065×10^{-4}		—	—	—	—	—	—	—	—
0.05	3.8882×10^{-3}	3.7821×10^{-3}	3.7786×10^{-3}		1.00390	1.00384	1.00384	0.00006	3.0039	3.0322	3.0322	0.0001
0.0769	0.033166	0.031171	0.031069		1.03316	1.03167	1.03167	0.0005	3.0332	3.0322	3.0322	0.0010
0.1	0.11638	0.10674	0.10601		1.1164	1.1083	1.1080	0.0016	3.1164	3.1102	3.1102	0.0035
0.125	0.30968	0.28090	0.27807		1.3097	1.2841	1.2822	0.0032	3.3097	3.2886	3.2879	0.0077
0.167	0.91141	0.83202	0.82217		1.9114	1.8382	1.8301	0.0062	3.9114	3.8481	3.8423	0.0161
0.2	1.6084	1.4881	1.4721		2.6084	2.4964	2.4821	0.0083	4.6085	4.5098	4.4984	0.0217
0.25	2.8797	2.7192	2.6941		3.8797	3.7292	3.7060	0.0100	5.8796	5.7458	5.7258	0.0266
0.333	5.2554	5.0847	5.0536		6.2555	6.0944	6.0651	0.0097	8.2556	8.1110	8.0846	0.0263
0.5	10.0915	9.9746	9.9487		11.0915	10.9805	10.9557	0.0059	13.0916	12.9907	12.9679	0.0161
0.667	14.747	14.677	14.660		15.748	15.680	15.663	0.003	17.747	17.686	17.670	0.0091
0.833	19.253	19.210	19.198		20.253	20.211	—	0.001	22.253	22.214	—	0.004
1	23.662	23.634	23.626		24.662	24.635	—	0.001	26.662	26.637	—	0.003
1.25	30.165	30.150	30.145		31.165	31.150	—	0.000	33.165	33.151	—	0.001

An idea of the magnitude of the correction terms, and of the rapidity with which the series (49) converges, can be obtained by an inspection of table 1. It is seen that the deviations from the zero-order values of the resistance are negative (as noted in § 3.31 above), are largest in relative value when T/Θ is of the order of 0.2 but decrease rapidly for higher and lower temperatures, and are most important for very pure specimens, the maximum value of the correction to the ideal resistance amounting to about 10 %. The deviations Δ from Matthiessen's rule are positive, increase with the amount of impurity present and are largest in the region where the residual and ideal resistances are of the same order of magnitude. They are, however, very small in relative magnitude, and, for the range of values shown in table 1, Δ is always less than 1 % of the total resistance.

The measured resistance values are usually separated into an ideal and a residual part by assuming the validity of Matthiessen's rule, and the temperature variation of the ideal resistance is assumed to be given by the Grüneisen-Bloch formula (35) (Grüneisen 1933). It is now seen that both these assumptions are incorrect, and in particular that equation (35) leads to values of the ideal resistance which are considerably too high at intermediate temperatures. Also Matthiessen's rule is strictly valid only at high temperatures, and at very low temperatures such that the residual resistance is large compared with the ideal resistance; however, the rule always represents a good approximation, and the deviations from it need only be taken into account if high accuracy is required. The experimental data (compare the tables given by Grüneisen 1933) are not sufficiently accurate to be compared in detail with the theoretical results, but they are in general agreement with the predictions outlined above. The observed values of the ideal resistance at intermediate temperatures are in most cases smaller than the values calculated from equation (35), and the magnitude of the deviations is of the order predicted by the present theory. The observed deviations from Matthiessen's rule, however, are in many cases too large to be explained by the present theory, although they agree qualitatively with its predictions; and to account for their order of magnitude it is, in general, necessary to employ models more complicated than the one-band free-electron model discussed here (Sondheimer & Wilson 1947, p. 450; Kohler 1949*b*).

4.22. Kohler (1949*b*) has proved generally that the deviations Δ from Matthiessen's rule are always positive (or zero), and it has been shown above that the behaviour of the free-electron model is in agreement with this result. A calculation due to Dube (1938), however, is in apparent contradiction with Kohler's theorem. Dube attempted to obtain an estimate of Δ for the free-electron model by using Wilson's (1937) method of successive approximations to calculate the resistance of an impure metal at low temperatures, correct to terms of order $(\rho_i/\rho_r)^2$. Dube's result is (using his notation)

$$\rho = \rho_r + \rho_i - \{g(T) - 1\} \rho_i^2 / \rho_r, \quad (52)$$

where ρ_i is given by (35), and where $g(T) - 1$ is a complicated function of T which tends to a constant positive value at very low temperatures. (An expression corresponding to (52) may easily be obtained from (44) by expanding in powers of σ_r/σ_i , but owing to the restricted validity of such an expansion this question will not be considered further here.)

Dube used his result to conclude that the deviation from Matthiessen's rule is

$$\Delta = \rho - (\rho_r + \rho_i) = -\{g(T) - 1\} \rho_i^2 / \rho_r, \quad (53)$$

which is negative. This argument is, however, fallacious. The reason is that, in the approximation considered by Dube, it is *not* correct to identify the expression (35) with the ideal resistance. In fact, (35) represents merely the leading term in an expansion for the true ideal resistance ρ'_i at low temperatures, of the form

$$\rho'_i = \rho_i \left[1 - C \left(\frac{T}{\Theta} \right)^2 + O \left(\left(\frac{T}{\Theta} \right)^4 \right) \right], \quad (54)$$

where C is a positive constant. The leading term in the expression for Δ at low temperatures is therefore given by

$$\Delta = \rho - (\rho_r + \rho'_i) = C \rho_i \left(\frac{T}{\Theta} \right)^2, \quad (55)$$

and not by Dube's expression (53), the latter being of higher order in T/Θ . Δ is therefore positive, in agreement with the general result, and it is clear that Dube's calculation by itself cannot be used to draw any conclusions whatsoever as to the nature of the deviations from Matthiessen's rule.

5. SECOND-ORDER PHENOMENA

5.1. The thermal conductivity

The general expressions (39) and (45) for the thermal conductivity are analogous to those for the electrical conductivity, and may be discussed in similar fashion. The zero-order approximation $\kappa^{(0)}$ given by equations (37) and (46) was first obtained by Wilson (1937) and has been discussed in detail by Makinson (1938), and we must now consider how this expression has to be modified to take into account the effect of the higher order correction terms. Wilson has shown that $\kappa^{(0)}$ is the correct expression for the thermal conductivity at high temperatures, including terms of order $(\Theta/T)^2$ and $M\Theta/T$, and that it holds also for an impure metal at low temperatures, correct to terms of order κ_r/κ_i . There is, however, one important difference compared with the case of the electrical conductivity. Bloch's (1930) method, which is used to show that the ideal electrical resistance at low temperatures is given correctly by (35), cannot be applied to the second-order effects such as the thermal conductivity, and in fact the ideal thermal conductivity at low temperatures is *not* given by (37), contrary to statements made by Kroll (1933*b*) and Wilson (1936, p. 219). This matter is discussed further below, but the difference in status between equations (35) and (37) should be carefully noted.

5.11. To evaluate the thermal conductivity numerically, we use the notation of § 4.2 to write equation (45) in a form which corresponds to (49). This is

$$\frac{12A}{\pi^2 L_0 \Theta} \kappa = \frac{T/\Theta}{\left(\frac{T}{\Theta} \right)^5 F_{11} + a_2 \frac{\rho_r}{4A}} + \frac{9}{\pi^4 \Theta} T \sum_{n=1}^{\infty} \frac{(\chi_a^{(2n)})^2}{\chi^{(2n-2)} \chi^{(2n)}}, \quad (56)$$

where A is given by (48), $L_0 = \frac{1}{3}(\pi k/\epsilon)^2$ is the 'normal' (high-temperature) value of the Wiedemann-Franz ratio, and where

$$\chi^{(2n)} = \begin{vmatrix} \left(\frac{T}{\Theta}\right)^5 F_{11} + a_2 \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{13} + a_4 \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{1, 2n+1} + a_{2n+2} \frac{\rho_r}{4A} \\ \left(\frac{T}{\Theta}\right)^5 F_{31} + a_4 \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{33} + a_6 \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{3, 2n+1} + a_{2n+4} \frac{\rho_r}{4A} \\ \dots & \dots & \dots & \dots \\ \left(\frac{T}{\Theta}\right)^5 F_{2n+1, 1} + a_{2n+2} \frac{\rho_r}{4A} & \left(\frac{T}{\Theta}\right)^5 F_{2n+1, 3} + a_{2n+4} \frac{\rho_r}{4A} & \dots & \left(\frac{T}{\Theta}\right)^5 F_{2n+1, 2n+1} + a_{4n+2} \frac{\rho_r}{4A} \end{vmatrix}, \quad (57)$$

while $\chi_a^{(2n)}$ is obtained from $\chi^{(2n)}$ on replacing the last column by $a_2, a_4, \dots, a_{2n+2}$. The value of the thermal conductivity obtained by retaining N terms of the infinite series in (56) will be denoted by $\kappa^{(2N)}$.* The $F_{\mu\nu}$'s required for evaluating the first three approximations are given explicitly in appendix 2; the results of the calculations for a monovalent metal, both in the absence of impurities and for two values of the impurity parameter $\rho_r/4A$, are shown in table 2 and are displayed graphically in figures 1 to 3.

As a particular result it should be noted that, for an ideal metal at low temperatures, *each* term of the series (56) is proportional to $(\Theta/T)^2$; the ideal thermal conductivity is therefore proportional to T^{-2} , as stated by Makinson, but the proportionality constant cannot be given in closed form. (It may be obtained from equation (39) as the ratio of two infinite determinants.) In the presence of impurities, however, the zero-order approximation is exact at sufficiently low temperatures (as stated above), being proportional to $(T/\Theta)/(\rho_r/4A)$, whereas each correction term is proportional to $(T/\Theta)^7/(\rho_r/4A)^3$. At high temperatures $\kappa^{(0)}$ is independent of T , and each correction term (for a pure metal) is proportional to $(\Theta/T)^4$.

The zero-order curves in figures 1 to 3 correspond to Makinson's approximation (compare figure 1 of Makinson's paper). The present theory shows that Makinson's formula for the thermal conductivity is qualitatively correct, and in particular it confirms the existence of a minimum in κ at intermediate temperatures, provided that the amount of impurity is not too high.† The numerical values of the thermal conductivity, however, are increased at intermediate temperatures in all cases, and are increased particularly at low temperatures in the case of an ideally pure metal. The position of the minimum in κ is shifted towards slightly higher temperatures, and the same is true for the low-temperature maximum in κ which occurs for impure metals. The correction terms are, in general, larger than in the case of the electrical conductivity, the first correction to the ideal thermal conductivity alone amounting to about 25 % at low temperatures. This means that the Lorenz

* The first two terms of the series for the case of an ideally pure metal have previously been obtained and discussed by Kroll (1938) and by Umeda & Yamamoto (1949).

† It may be objected that inclusion of the higher approximations might invalidate this result. This is not so, since the thermal conductivity at high temperatures must decrease at first as the temperature is lowered (Wilson 1937, p. 376).

TABLE 2. THE THERMAL CONDUCTIVITY OF MONOVALENT METALS

T/Θ	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(0)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(2)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(4)}$	$\rho_i/4A = 0$			$\rho_i/4A = 0.01$			$\rho_i/4A = 0.03$		
	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(0)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(2)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa_i^{(4)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(0)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(2)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(4)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(0)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(2)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(4)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(0)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(2)}$	$\frac{12A}{\pi^2 L_0 \Theta} \kappa^{(4)}$
≤ 1	$0.012757 \left(\frac{\Theta}{T}\right)^2$	$0.015705 \left(\frac{\Theta}{T}\right)^2$	$0.016957 \left(\frac{\Theta}{T}\right)^2$	$30.396 \frac{T}{\Theta}$	$30.396 \frac{T}{\Theta}$	$30.396 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$	$10.132 \frac{T}{\Theta}$
0.05	5.175	6.404	6.835	1.1748	1.2086	1.2117	0.4614	0.4644	0.4644	0.4614	0.4644	0.4644
0.0769	2.246	2.807	3.036	1.1457	1.2548	1.2799	0.5786	0.5990	0.6013	0.5786	0.5990	0.6013
0.1	1.4180	1.7552	1.8872	0.9669	1.0972	1.1348	0.5909	0.6294	0.6370	0.5909	0.6294	0.6370
0.125	1.0347	1.2362	1.3132	0.8132	0.9229	0.9585	0.5695	0.6140	0.6257	0.5695	0.6140	0.6257
0.167	0.8005	0.8912	0.9265	0.6913	0.7534	0.7762	0.5431	0.5763	0.5874	0.5431	0.5763	0.5874
0.2	0.7446	0.7955	0.8162	0.6634	0.7009	0.7158	0.5445	0.5669	0.5753	0.5445	0.5669	0.5753
0.25	0.7372	0.7604	0.7708	0.6720	0.6901	0.6982	0.5710	0.5827	0.5878	0.5710	0.5827	0.5878
0.333	0.7909	0.7983	0.8023	0.7336	0.7397	0.7429	0.6408	0.6450	0.6472	0.6408	0.6450	0.6472
0.5	0.9214	0.9225	—	0.8687	0.8697	—	0.7796	0.7803	—	0.7796	0.7803	—
0.667	1.0133	1.0136	—	0.9651	—	—	0.8811	—	—	0.8811	—	—
0.833	1.0718	1.0718	—	1.0282	—	—	0.9510	—	—	0.9510	—	—
1	1.1094	—	—	1.0703	—	—	0.9999	—	—	0.9999	—	—
1.25	1.1440	—	—	1.1106	—	—	1.0492	—	—	1.0492	—	—
≥ 1	1.2158	—	—	1.2158	—	—	1.2158	—	—	1.2158	—	—

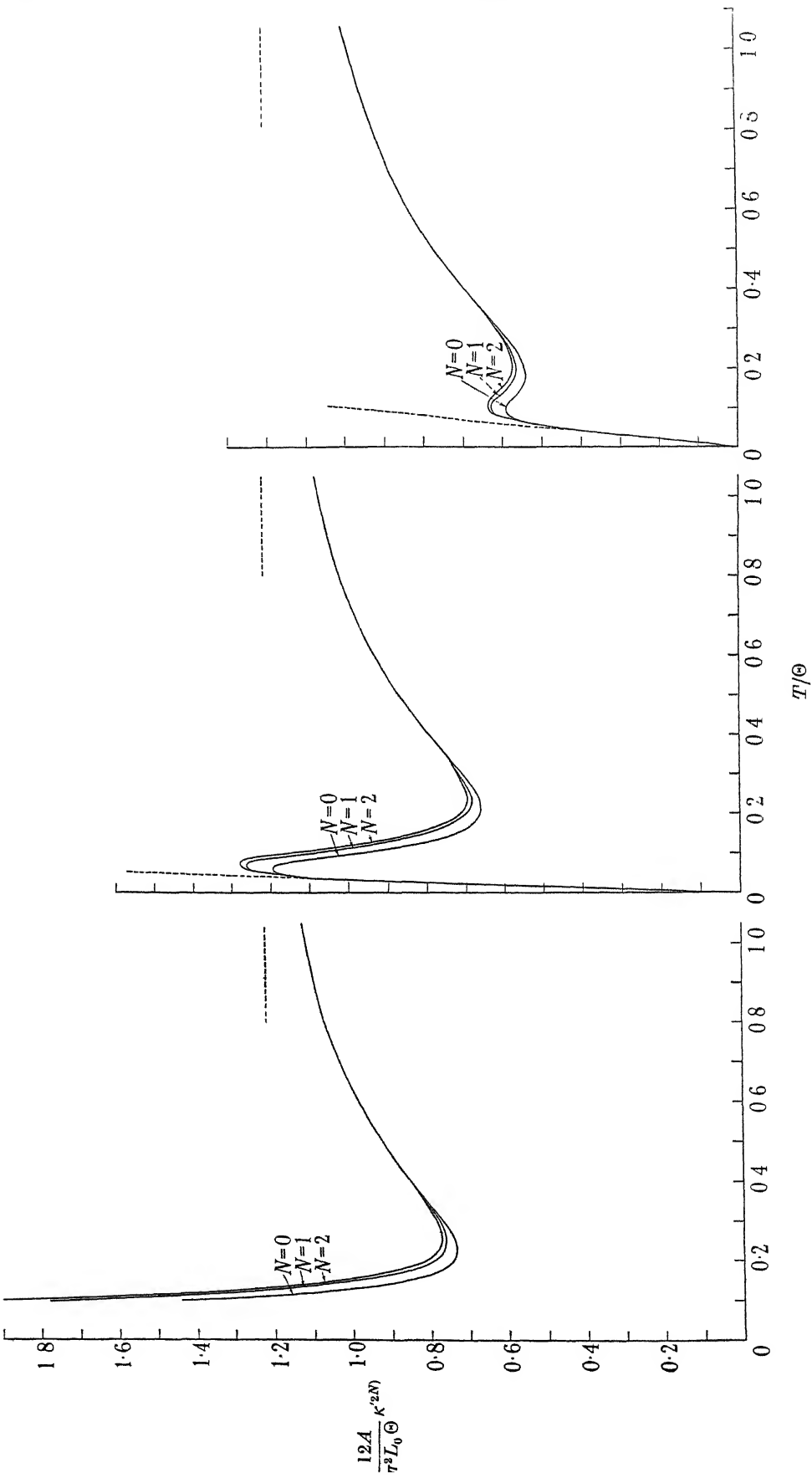


FIGURE 1. $\rho_r/4A = 0$

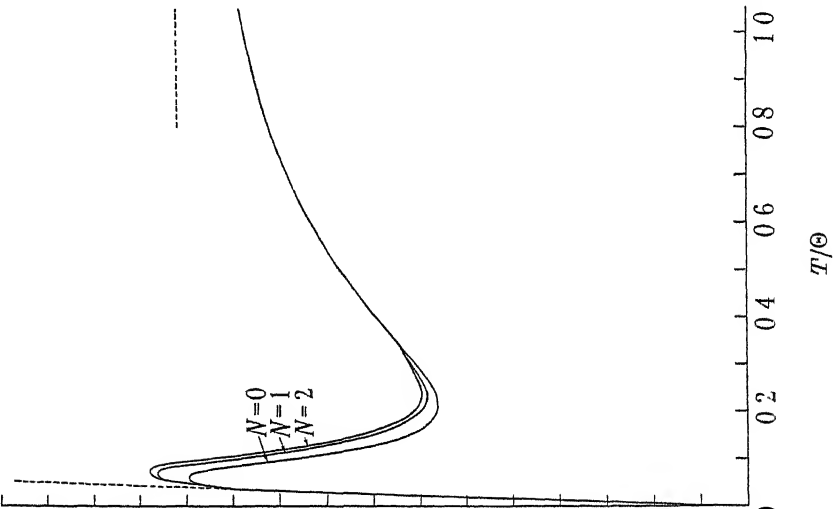


FIGURE 2. $\rho_r/4A = 0.01$

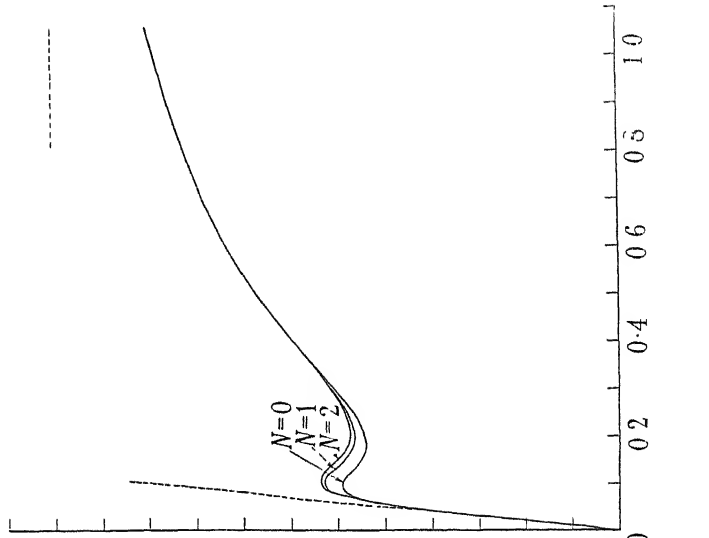


FIGURE 3. $\rho_r/4A = 0.03$

FIGURES 1 TO 3. The thermal conductivity of monovalent metals.

number $L = \kappa/\sigma T$ is increased above Makinson's values; numerical estimates of L may be obtained by combining corresponding values of σ and κ given in tables 1 and 2.

The experimental results on the thermal conductivity have been discussed by Makinson; they are not sufficiently accurate to justify renewed discussion in the light of the present theory. There is fair agreement between theory and experiment, both qualitatively and as regards orders of magnitude, but the minimum in κ predicted by the theory does not seem to have been observed.

5.2. *The thermo-electric power*

The general expression (40) for the thermo-electric power is more complicated in character than the corresponding expressions for the electrical and thermal conductivities, and will not be discussed in the same detail. It is known that, at high temperatures,

$$\mathfrak{S} = \pi^2 k^2 T / \zeta, \quad (58)$$

and that, at very low temperatures,

$$\mathfrak{S} = \pi^2 k^2 T / 3\zeta, \quad (59)$$

the latter result, in particular, being valid both for an ideal metal and in the presence of impurities (Kroll 1933*a*; Wilson 1936, pp. 177, 220). The correction terms of order $(\Theta/T)^2$ at high temperatures, and of order σ_i/σ for an impure metal at low temperatures, have been obtained by Wilson (1937). These results may be shown to follow as special cases from equation (40), which provides the connexion formula for the transition from the expression (59) at low temperatures to the expression (58) at high temperatures, and which replaces the simple approximate interpolation formulae given by Kohler (1949*a*) and Sondheimer (1947). Kohler's formula, in particular, is obtained if the infinite determinants in (40) are broken off at the third row and column, and higher approximations to \mathfrak{S} may be obtained by retaining additional rows and columns. It does not seem possible, however, to make any general statements regarding the effect on \mathfrak{S} of retaining the higher terms. There is therefore no point in discussing them in detail, particularly in view of the well-known inadequacy of the free-electron model to serve as a basis for discussing the observed thermo-electric effects in real metals.

The above work was carried out while the author was spending a year in the United States as a member of the Research Laboratory of Electronics of the Massachusetts Institute of Technology, and he wishes to thank Professors J. C. Slater and A. G. Hill for enabling him to come to the Institute and for the hospitality extended to him during his stay there. He also wishes to thank Mr A. H. Wilson, F.R.S., for his interest in the problem and for reading the manuscript, Mr S. Sensiper for several valuable discussions, and Miss Elizabeth J. Campbell for performing most of the numerical work.

The work has been supported in part by the U.S. Signal Corps, the Office of Naval Research and the Air Materiel Command.

APPENDIX 1. EVALUATION OF $\alpha_{\mu}^{(n)}$ AND $d_{\mu\nu}$

A 1.1. Consider first
$$\alpha_{\mu}^{(n)} = \int_{-\infty}^{\infty} E^n \eta^{\mu} \frac{\partial f_0}{\partial \eta} d\eta, \quad (\text{A } 1)$$

where $\eta = (E - \zeta)/kT$. Put $\gamma = kT/\zeta$, and suppose that E^n can be expanded in the series

$$E^n = \zeta^n (1 + \gamma\eta)^n = \zeta^n + \zeta^n \sum_{r=1}^{\infty} \frac{n(n-1)(n-2)\dots(n-r+1)}{r!} (\gamma\eta)^r, \quad (\text{A } 2)$$

so that (assuming that the series can be integrated term by term)

$$\alpha_{\mu}^{(n)} = \zeta^n \int_{-\infty}^{\infty} \eta^{\mu} \frac{\partial f_0}{\partial \eta} d\eta + \zeta^n \sum_{r=1}^{\infty} \frac{n(n-1)(n-2)\dots(n-r+1)}{r!} \gamma^r \int_{-\infty}^{\infty} \eta^{r+\mu} \frac{\partial f_0}{\partial \eta} d\eta. \quad (\text{A } 3)$$

Now (Wilson 1936, appendix 1.3)

$$\left. \begin{aligned} \int_{-\infty}^{\infty} \eta^r \frac{\partial f_0}{\partial \eta} d\eta &= - \int_{-\infty}^{\infty} \frac{\eta^r e^{-\eta}}{(1 + e^{-\eta})^2} d\eta = -a_r \quad (r \text{ even}), \\ &= 0 \quad (r \text{ odd}), \end{aligned} \right\} \quad (\text{A } 4)$$

where

$$\left. \begin{aligned} a_0 &= 1, \\ a_{2s} &= 2(2s)! \sum_{l=1}^{\infty} \frac{(-1)^{l+1}}{l^{2s}} \quad (s \geq 1). \end{aligned} \right\} \quad (\text{A } 5)$$

Combining (A 3) and (A 4), we obtain equations (24a) and (24b) of the main text.

A 1.2. Next, combining equations (3) and (11) of the main text and putting $E = \zeta(1 + \gamma\eta)$, it is found that

$$d_{\mu\nu} = -\zeta \int_0^{\Theta/T} \frac{z^2 dz}{e^z - 1} I_{\mu\nu}(z), \quad (\text{A } 6)$$

where

$$\begin{aligned} I_{\mu\nu}(z) = \int_{-\infty}^{\infty} \eta^{\mu} \left[\left\{ (1 + \gamma\eta) \eta^{\nu} - (\eta + z)^{\nu} \left(1 + \gamma\eta + \frac{1}{2}\gamma z - \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 \right) \right\} \frac{e^{\eta} + 1}{e^{\eta} + e^{-z}} \right. \\ \left. + \left\{ (1 + \gamma\eta) \eta^{\nu} - (\eta - z)^{\nu} \left(1 + \gamma\eta - \frac{1}{2}\gamma z - \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 \right) \right\} \frac{e^{\eta} + 1}{e^{\eta} - z + 1} \right] \frac{\partial f_0}{\partial \eta} d\eta, \end{aligned} \quad (\text{A } 7)$$

the term corresponding to the presence of impurities having been omitted.

Detailed consideration will only be given to the case where μ and ν are both odd, the other cases being dealt with similarly. In (A 7), expand $(\eta \pm z)^{\nu}$ in powers of η , and put

$$J_s^{\pm}(z) = \int_{-\infty}^{\infty} \eta^s \left(\frac{e^{\eta} + 1}{e^{\eta} + e^{-z}} \pm \frac{e^{\eta} + 1}{e^{\eta} - z + 1} \right) \frac{\partial f_0}{\partial \eta} d\eta. \quad (\text{A } 8)$$

This gives, after some rearrangement,

$$\begin{aligned} I_{\mu\nu} = z^{\nu} \left\{ \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right\} J_{\mu}^{-} - \frac{1}{2} \gamma z^{\nu+1} J_{\mu}^{+} + \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 J_{\mu+\nu}^{+} - (\nu + \frac{1}{2}) \gamma z J_{\mu+\nu}^{-} \\ + \sum_{r=2,4,6,\dots}^{\nu-1} \left[\binom{\nu}{r} \left\{ \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right\} J_{\mu+\nu-r}^{+} - \gamma z \left(\frac{1}{2} \binom{\nu}{r} + \binom{\nu}{r+1} \right) J_{\mu+\nu-r}^{-} \right] z^r \\ + \sum_{r=1,3,5,\dots}^{\nu-2} \left[\binom{\nu}{r} \left\{ \frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right\} J_{\mu+\nu-r}^{-} - \gamma z \left(\frac{1}{2} \binom{\nu}{r} + \binom{\nu}{r+1} \right) J_{\mu+\nu-r}^{+} \right] z^r. \end{aligned} \quad (\text{A } 9)$$

A1.21 To evaluate the integrals $J_s^\pm(z)$ use is made of the following result, which is a slight generalization of a formula given by Wilson (1937, equation (11)):

$$\int_{-\infty}^{\infty} \frac{\eta^s}{(e^{-\eta} + 1)(e^{\eta+z} + 1)} d\eta = \frac{1}{(s+1)(e^z - 1)} \int_{-\infty}^{\infty} \frac{\eta^{s+1} - (\eta - z)^{s+1}}{(e^{\eta} + 1)(e^{-\eta} + 1)} d\eta. \quad (\text{A } 10)$$

This gives $J_s^+(z) = \frac{1}{(s+1)(1-e^{-z})} \int_{-\infty}^{\infty} \{(\eta+z)^{s+1} - (\eta-z)^{s+1}\} \frac{\partial f_0}{\partial \eta} d\eta,$ (A 11a)

and $J_s^-(z) = \frac{1}{(s+1)(1-e^{-z})} \int_{-\infty}^{\infty} \{2\eta^{s+1} - (\eta+z)^{s+1} - (\eta-z)^{s+1}\} \frac{\partial f_0}{\partial \eta} d\eta.$ (A 11b)

Expanding $(\eta \pm z)^{s+1}$ in powers of η and using (A 4), it is found that

$$J_s^+(z) = -\frac{2}{(s+1)(1-e^{-z})} \sum_{t=1,3,5,\dots}^{s+1} \binom{s+1}{t} a_{s-t+1} z^t, \quad J_s^-(z) = 0 \quad (\text{A } 12a)$$

if s is even, and

$$J_s^+(z) = 0, \quad J_s^-(z) = \frac{2}{(s+1)(1-e^{-z})} \sum_{t=2,4,6,\dots}^{s+1} \binom{s+1}{t} a_{s-t+1} z^t \quad (\text{A } 12b)$$

if s is odd.

A1.22. From (A 9) and (A 12), we now obtain

$$\begin{aligned} I_{\mu\nu} = & \frac{2z^\nu}{(\mu+1)(1-e^{-z})} \left\{ \frac{D}{\xi} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right\} \sum_{t=2,4,6,\dots}^{\mu+1} \binom{\mu+1}{t} a_{\mu-t+1} z^t \\ & - \frac{2}{(\mu+\nu+1)(1-e^{-z})} \frac{D}{\xi} \left(\frac{T}{\Theta} \right)^2 z^2 \sum_{t=1,3,5,\dots}^{\mu+\nu+1} \binom{\mu+\nu+1}{t} a_{\mu+\nu-t+1} z^t \\ & - \frac{2}{1-e^{-z}} \left\{ \frac{D}{\xi} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right\} \left[\sum_{r=2,4,6,\dots}^{\nu-1} \frac{1}{\mu+\nu-r+1} \binom{\nu}{r} z^r \sum_{t=1,3,5,\dots}^{\mu+\nu-r+1} \binom{\mu+\nu-r+1}{t} \right. \\ & \left. \times a_{\mu+\nu-r-t+1} z^t - \sum_{r=1,3,5,\dots}^{\nu-2} \frac{1}{\mu+\nu-r+1} \binom{\nu}{r} z^r \sum_{t=2,4,6,\dots}^{\mu+\nu-r+1} \binom{\mu+\nu-r+1}{t} a_{\mu+\nu-r-t+1} z^t \right]. \end{aligned} \quad (\text{A } 13)$$

The double sums in (A 13) may be simplified by collecting together all the terms involving a fixed power of z , say z^k . The square bracket in (A 13) then becomes

$$\sum_{k=3,5,7,\dots}^{\nu} A_{\mu\nu,k} \frac{a_{\mu+\nu-k+1}}{(\mu+\nu-k+1)!} z^k + \sum_{k=\nu+2,\nu+4,\dots}^{\mu+\nu+1} B_{\mu\nu,k} \frac{a_{\mu+\nu-k+1}}{(\mu+\nu-k+1)!} z^k, \quad (\text{A } 14)$$

where $A_{\mu\nu,k} = \sum_{r=1}^{k-1} (-1)^r \binom{\nu}{r} \frac{(\mu+\nu-r)!}{(k-r)!},$ (A 15a)

and $B_{\mu\nu,k} = \sum_{r=1}^{\nu-1} (-1)^r \binom{\nu}{r} \frac{(\mu+\nu-r)!}{(k-r)!}.$ (A 15b)

The summations in (A 15) may be evaluated to give (supposing that $\mu \leq \nu$)

$$A_{\mu\nu,k} = (\mu+\nu-k)! \left\{ \binom{\mu}{k} + \binom{\nu}{k} \right\} - \frac{(\mu+\nu)!}{k!} \quad (\text{A } 16a)$$

if $k \leq \mu,$ $A_{\mu\nu,k} = (\mu+\nu-k)! \binom{\nu}{k} - \frac{(\mu+\nu)!}{k!}$ (A 16b)

$$\text{if } \mu + 2 \leq k \leq \nu, \quad B_{\mu\nu, k} = \frac{\mu!}{(k-\nu)!} - \frac{(\mu+\nu)!}{k!} \quad (\text{A } 16c)$$

if $\nu + 2 \leq k \leq \mu + \nu - 1$, and

$$B_{\mu\nu, \mu+\nu+1} = \frac{1}{\mu+1} - \frac{1}{\mu+\nu+1} - \frac{\mu! \nu!}{(\mu+\nu+1)!}. \quad (\text{A } 16d)$$

Combining (A 13), (A 14) and (A 16) and simplifying, we obtain

$$\begin{aligned} I_{\mu\nu} = & -\frac{2}{1-e^{-z}} \left[\frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^3 a_{\mu+\nu} + \sum_{k=3,5,7,\dots}^{\mu+\nu+1} \binom{\mu+\nu+1}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu+1} z^k \right. \\ & + \left. \left(\frac{D}{\zeta} \left(\frac{T}{\Theta} \right)^2 z^2 - 1 \right) \left\{ \sum_{k=3,5,7,\dots}^{\mu} \binom{\mu}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} z^k + \sum_{k=3,5,7,\dots}^{\nu} \binom{\nu}{k} \frac{a_{\mu+\nu-k+1}}{\mu+\nu-k+1} z^k \right. \right. \\ & \left. \left. - \frac{\mu! \nu!}{(\mu+\nu+1)!} z^{\mu+\nu+1} \right\} \right]. \quad (\text{A } 17) \end{aligned}$$

The same result is obtained if $\mu > \nu$, and combination of (A 6) and (A 17) now leads to equations (26a) and (27a) of the main text. The impurity term has here been omitted, it is easily evaluated by the method used in appendix 1.1 above, and the details will not be given.

APPENDIX 2. NUMERICAL EVALUATION OF THE TRANSPORT MAGNITUDES

A 2.1. Explicit expressions are given below for the a_μ 's and $F_{\mu\nu}$'s required for evaluating the first three approximations to the electrical and thermal conductivities according to equations (49) and (56). Equation (25) gives

$$a_0 = 1, \quad a_2 = \frac{1}{3}\pi^2, \quad a_4 = \frac{7}{15}\pi^4, \quad a_6 = \frac{31}{21}\pi^6, \quad a_8 = \frac{127}{15}\pi^8, \quad a_{10} = \frac{2555}{33}\pi^{10}. \quad (\text{A } 18)$$

Further, equations (27a) and (28a) give, for μ and ν even,

$$F_{00} = \mathcal{J}_5, \quad (\text{A } 19)$$

$$F_{02} = \frac{1}{3}\pi^2 \mathcal{J}_5 + \frac{1}{3}\mathcal{J}_7, \quad (\text{A } 20)$$

$$F_{22} = \frac{7}{15}\pi^4 \mathcal{J}_5 + \frac{1}{30}\mathcal{J}_9 + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{2}{3}\pi^2 \mathcal{J}_5 + \frac{1}{6}\mathcal{J}_7 \right), \quad (\text{A } 21)$$

$$F_{04} = \frac{7}{15}\pi^4 \mathcal{J}_5 + \frac{2}{3}\pi^2 \mathcal{J}_7 + \frac{1}{5}\mathcal{J}_9, \quad (\text{A } 22)$$

$$F_{24} = \frac{31}{21}\pi^6 \mathcal{J}_5 + \frac{7}{15}\pi^4 \mathcal{J}_7 + \frac{1}{105}\mathcal{J}_{11} + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{25}{15}\pi^4 \mathcal{J}_5 + \pi^2 \mathcal{J}_7 + \frac{2}{15}\mathcal{J}_9 \right), \quad (\text{A } 23)$$

$$\begin{aligned} F_{44} = & \frac{127}{15}\pi^8 \mathcal{J}_5 + \frac{124}{63}\pi^6 \mathcal{J}_7 + \frac{1}{630}\mathcal{J}_{13} \\ & + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{248}{21}\pi^6 \mathcal{J}_5 + \frac{98}{15}\pi^4 \mathcal{J}_7 + \frac{4}{3}\pi^2 \mathcal{J}_9 + \frac{23}{210}\mathcal{J}_{11} \right), \quad (\text{A } 24) \end{aligned}$$

and, for μ and ν odd,

$$F_{11} = \frac{1}{3}\pi^2 \mathcal{J}_5 - \frac{1}{6}\mathcal{J}_7 + \frac{1}{2} \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \mathcal{J}_5, \quad (\text{A } 25)$$

$$F_{13} = \frac{7}{15}\pi^4 \mathcal{J}_5 + \frac{1}{6}\pi^2 \mathcal{J}_7 - \frac{1}{20}\mathcal{J}_9 + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{1}{2}\pi^2 \mathcal{J}_5 + \frac{1}{4}\mathcal{J}_7 \right), \quad (\text{A } 26)$$

$$F_{33} = \frac{31}{21}\pi^6 \mathcal{J}_5 + \frac{7}{30}\pi^4 \mathcal{J}_7 - \frac{1}{140}\mathcal{J}_{11} + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{21}{10}\pi^4 \mathcal{J}_5 + \pi^2 \mathcal{J}_7 + \frac{3}{20}\mathcal{J}_9 \right), \quad (\text{A } 27)$$

$$F_{15} = \frac{3}{21}\pi^6 \mathcal{J}_5 + \frac{7}{6}\pi^4 \mathcal{J}_7 + \frac{1}{6}\pi^2 \mathcal{J}_9 - \frac{1}{42}\mathcal{J}_{11} + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{7}{6}\pi^4 \mathcal{J}_5 + \frac{5}{6}\pi^2 \mathcal{J}_7 + \frac{1}{6}\mathcal{J}_9 \right), \quad (\text{A } 28)$$

$$F_{35} = \frac{1}{15}\pi^8 \mathcal{J}_5 + \frac{3}{126}\pi^6 \mathcal{J}_7 + \frac{7}{60}\pi^4 \mathcal{J}_9 - \frac{1}{504}\mathcal{J}_{13} + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{15}{14}\pi^6 \mathcal{J}_5 + \frac{7}{12}\pi^4 \mathcal{J}_7 + \frac{4}{3}\pi^2 \mathcal{J}_9 + \frac{1}{168}\mathcal{J}_{11} \right), \quad (\text{A } 29)$$

$$F_{55} = \frac{2}{33}\pi^{10} \mathcal{J}_5 + \frac{1}{6}\pi^8 \mathcal{J}_7 + \frac{3}{63}\pi^6 \mathcal{J}_9 - \frac{1}{2772}\mathcal{J}_{15} + \frac{\zeta}{D} \left(\frac{\Theta}{T} \right)^2 \left(\frac{63}{6}\pi^8 \mathcal{J}_5 + \frac{38}{63}\pi^6 \mathcal{J}_7 + 14\pi^4 \mathcal{J}_9 + \frac{5}{3}\pi^2 \mathcal{J}_{11} + \frac{2}{252}\mathcal{J}_{13} \right). \quad (\text{A } 30)$$

A 2.2. The integrals

$$\mathcal{J}_n(x) = \int_0^x \frac{z^n dz}{(e^z - 1)(1 - e^{-z})} \quad (\text{A } 31)$$

(with $x = \Theta/T$) are most conveniently evaluated by means of the series

$$\mathcal{J}_n(x) = n! \sum_{s=1}^{\infty} \frac{1}{s^n} - \frac{x^n}{e^x - 1} - n! \sum_{s=1}^{\infty} \frac{e^{-sx}}{s^n} \left\{ 1 + sx + \frac{(sx)^2}{2!} + \dots + \frac{(sx)^{n-1}}{(n-1)!} \right\}, \quad (\text{A } 32)$$

valid for large x , and

$$\mathcal{J}_n(x) = x^{n-1} \left\{ \frac{1}{n-1} - \sum_{s=1}^{\infty} (-1)^{s-1} \frac{B_s}{(2s)!} \frac{2s-1}{2s+n-1} x^{2s} \right\}, \quad (\text{A } 33)$$

valid for small x , where the B_s are the Bernoulli numbers. Numerical values are given in table 3, correct to five significant figures (a table of values of \mathcal{J}_5 only has previously been given by Gruneisen 1933).

TABLE 3. VALUES OF $\mathcal{J}_n(\Theta/T)$

$T/\Theta \backslash n$	5	7	9	11	13	15	17
0	124.43	5082.1	3.6361×10^5	3.9937×10^7	6.2278×10^9	1.3077×10^{12}	3.5569×10^{14}
0.05	124.42	5078.2	3.6180×10^5	3.9083×10^7	5.8160×10^9	1.1030×10^{12}	2.5004×10^{14}
0.076923	123.14	4809.8	3.0344×10^5	2.5647×10^7	2.6594×10^9	3.0916×10^{11}	3.8962×10^{13}
0.1	116.38	3972.1	1.9743×10^5	1.2123×10^7	8.4470×10^8	6.3770×10^{10}	5.0803×10^{12}
0.125	101.48	2798.8	1.0353×10^5	4.4849×10^6	2.1340×10^8	1.0785×10^{10}	5.6799×10^{11}
0.16667	70.873	1328.9	31011	8.1272×10^5	2.2840×10^7	6.7201×10^8	2.0417×10^{10}
0.2	50.263	705.56	11953	2.2377×10^5	4.4520×10^6	9.2226×10^7	1.9661×10^9
0.25	29.488	281.75	3167.4	38845	5.0260×10^5	6.7413×10^6	9.2781×10^7
0.33333	12.771	72.010	468.43	3291.4	24264	1.8479×10^5	1.4409×10^6
0.5	3.2293	8.3763	24.717	78.230	258.74	881.95	3072.9
0.66667	1.1199	1.6538	2.7649	4.9460	9.2329	17.748	34.851
0.83333	0.47907	0.45534	0.48883	0.56090	0.67116	0.82665	1.0399
1	0.23662	0.15665	0.11700	0.093343	0.077632	0.066444	0.058071
1.25	0.098845	0.041987	0.020100	0.010273	0.0054720	0.0029989	0.0016782

REFERENCES

- Bloch, F. 1930 *Z. Phys.* **59**, 208.
 Dube, G. P. 1938 *Proc. Camb. Phil. Soc.* **34**, 559.
 Gruneisen, E. 1933 *Ann. Phys., Lpz.* (5), **16**, 530.
 Kohler, M. 1948 *Z. Phys.* **124**, 772.
 Kohler, M. 1949a *Z. Phys.* **125**, 679.

- Kohler, M. 1949^b *Z. Phys.* **126**, 495.
 Kowalewski, G. 1909 *Einführung in die Determinantentheorie*, § 41. Leipzig. Verlag von Veit und Comp.
 Kroll, W. 1933^a *Z. Phys.* **80**, 50.
 Kroll, W. 1933^b *Z. Phys.* **81**, 425.
 Kroll, W. 1938 *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **34**, 194.
 Makinson, R. E. B. 1938 *Proc. Camb. Phil. Soc.* **34**, 474.
 Sondheimer, E. H. 1947 *Proc. Camb. Phil. Soc.* **43**, 571.
 Sondheimer, E. H. & Wilson, A. H. 1947 *Proc. Roy. Soc. A*, **190**, 435.
 Umeda, K. 1942 *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **39**, 342.
 Umeda, K. & Toya, T. 1949 *J. Fac. Sci. Hokkaidō Univ.* (2), **3**, 257.
 Umeda, K. & Yamamoto, T. 1949 *J. Fac. Sci. Hokkaidō Univ.* (2), **3**, 249.
 Wilson, A. H. 1936 *The theory of metals*. Cambridge University Press.
 Wilson, A. H. 1937 *Proc. Camb. Phil. Soc.* **33**, 371.

The surface impedance of superconductors and normal metals at high frequencies

IV. Impedance at 9400 Mc./sec. of single crystals of normal and superconducting tin

By A. B. PIPPARD, *The Royal Society Mond Laboratory, University of Cambridge*

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 16 March 1950)

The measurements described in the earlier papers of this series have been extended to 9400 Mc./sec., a resonance technique being employed to determine the surface resistance of normal and superconducting tin, and the difference between the surface reactances of the material in the two states. Measurements on single crystals of different orientations have brought to light a marked anisotropy of all these quantities, of a kind which shows clearly the non-tensorial nature of the fundamental equations relating the field vectors. The prediction of the theory of the anomalous skin effect in normal metals, that the surface resistance should vary with frequency as $\omega^{\frac{1}{2}}$, is confirmed. The temperature variation of the resistance and reactance of superconducting tin has been studied in detail for a number of specimens of different orientations, and it has been found that over certain ranges of temperature the shapes of corresponding curves for different specimens are similar, apart from scaling factors depending on the orientation, the values of these scaling factors are used to characterize the surface impedance of each orientation.

INTRODUCTION

In the first three parts of this series (Pippard 1947), which will be referred to as I, II and III, an account was given of measurements of the surface impedance of normal metals and superconductors at a frequency of 1200 Mc./sec., and an attempt was made to fit the results into the framework of existing theories. Since that time the work of Reuter & Sondheimer (1948) has placed the phenomenon of the anomalous skin effect in normal metals on a secure theoretical basis, and their theory has been extended by Maxwell, Marcus & Slater (1949) to cover one special model of a superconductor. Meanwhile experiments have proceeded with a view to extending the

previous results to a frequency of 9400 Mc /sec., corresponding to a free-space wavelength of 3.2 cm. Preliminary accounts of this work have already been published (Pippard 1948, 1949*a*), but more recent measurements have shown that the phenomena are more complicated than was thought originally, owing to the marked anisotropy of tin. It has therefore been found necessary to study single crystals in detail, and the results of this investigation will be presented here. The additional labour involved in the investigation of single crystals of different orientations has been such as to render it inexpedient to study more than one metal, but this disadvantage is largely offset by the increased body of information concerning that one metal, tin, which, as will be seen later, makes it possible to derive results of some theoretical significance from the experimental behaviour.

APPARATUS

The experimental techniques employed in this work were similar in principle to those described in I (measurement of resistance) and III (measurement of reactance), and no more will be done here than to give a general description of the apparatus, pointing out those features of technique where there is some departure from earlier practice. As before, the specimen under investigation forms the principal element of a resonant system which may be excited by means of an oscillator of variable frequency, and from which a fraction of the energy may be extracted and rectified.

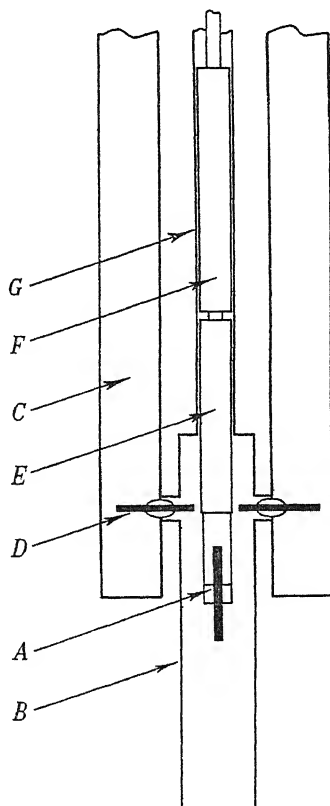


FIGURE 1. Schematic diagram of resonator.

The resonator assembly is shown diagrammatically in figure 1. The specimen, *A*, is a thin wire about $\frac{1}{2}\lambda$ long, λ being the wave-length, forming the inner conductor of a resonant co-axial line, of which the can, *B*, is the outer conductor; the diameter of the can was chosen to be not great enough to support any propagating wave-guide modes, and the length was sufficient to eliminate any significant end-effects. The length of the specimen then determines the resonant frequency; on account of the spreading of lines of force away from the ends of the specimen resonance occurs at a wave-length rather greater than twice the length of the specimen, as if there were a small capacitive load concentrated at each end. Power from the oscillator was brought to the resonator through a wave-guide, *C*, the actual coupling being effected by means of a probe, *D*. A similar probe and wave-guide extracted a fraction of the energy in the resonator and led it away to a detector. Since the resonator is to be immersed in liquid helium, while the oscillator and detector are at room temperature, it is necessary to minimize heat conduction along the wave-guides, and this was achieved by constructing them from german silver sheet of $\frac{1}{16}$ mm. thickness. They were then too weak to support even a small pressure difference between the inside and the outside, so that it was necessary to allow free access of liquid helium to the interior of the wave-guides by drilling suitable holes. On account of the low dielectric constant of liquid helium there was very little reflexion of power at the liquid surface, and no difficulties in measurement were occasioned by this arrangement. It was, however, essential to avoid the presence of liquid in the resonator, and even to take precautions to maintain constant the vapour density, since in the earlier measurements it was found that changes in vapour density caused troublesome changes of the resonant frequency. The probes, *D*, were therefore constructed from glass-metal seals, and the resonator was then a completely sealed system, except for valves at the top of the cryostat, by means of which the vapour pressure in the resonator could be adjusted and maintained at a suitable value.

As described in I, it is convenient, in order to determine the 'unloaded *Q*' of the resonator (that is, the value of *Q* when the coupling of the external circuits is so loose as to produce no effect on *Q*), to be able to vary the coupling between the specimen and the probes while keeping the system symmetrical. In the present arrangement the variation was simply effected by altering the height of the specimen in the can, and this alteration could be made from outside the cryostat. It was found important in preliminary experiments that the specimen should lie and move along the axis of the can; for this reason the specimen, mounted on a distrene rod, *E*, was attached by means of a ball-joint to a brass rod, *F*, which slid smoothly in the specimen tube, *G*. Before the experiment was begun the ball-joint was adjusted and clamped so that the specimen lay along the axis of the rod, and it was then found that on insertion into the resonator the coupling could be varied without destroying the symmetry of the system.

The specimen was mounted in such a way as to avoid as far as possible the presence of dielectric materials in regions where the electric field strength was high, especially near the ends of the specimen. The specimen holder may be seen in figure 1; the framework supporting the specimen is of fine silica tubes and threads, cemented together and to the distrene rod by means of distrene cement. It was estimated that the dielectric

loss with this arrangement should not limit the Q of the resonator to a lower value than 10^7 . Apart from this source of extraneous loss the only other source is the can, which may produce a loss comparable to that of the specimen. In all measurements of the resistance of superconducting tin the can was lined with lead by dipping; superconducting lead is by no means so good a conductor as tin at these frequencies, but on account of the large surface area of the can compared to that of the specimen it was only at the lowest temperatures that the loss in the can became as great as the loss in the specimen. The methods used for estimating the extraneous losses will be described when the experimental results are presented. In the measurements of normal resistance and of reactance the can was of brass, unlined, whose contribution to the loss was considerably smaller than that of the tin specimen, and could be allowed for with sufficient accuracy, so far as reactance measurements were concerned the loss in the can was of no consequence.

Outside the cryostat the principal pieces of equipment were the oscillator, wave-meters and detector. The oscillator was a standard CV 87 klystron, watercooled and protected against draughts, whose power supply was provided by a highly stabilized power pack. No attempt was made to provide further frequency stabilization, since by itself it was remarkably stable, drifting at the rate of about 1 part in 10^6 per min., and occasionally subject to sudden jumps in frequency of 1 part in 10^7 . Only when the value of Q exceeded 10^5 were these irregularities noticeable, and it was possible with a little care to measure values of 5×10^5 with an accuracy of about 5 %. In order to be able to vary the frequency over a small range, for the measurement of Q , a fine tuning adjustment was provided in the form of four silica rods, mounted on a geometrical slide so that they could be inserted into the rhumbatron. By this means as fine control was possible of the frequency as was demanded by the value of Q being measured. The overall range of tuning with the silica rods was 0.2 %, and over this range the power output varied only slightly.

To calibrate the oscillator two wavemeters were needed, the first a cavity wave-meter to determine the actual frequency within 0.1 %, the second a differential wavemeter for calibration of the fine adjustment. The heterodyne differential wavemeter described in I was found to be inconvenient at this frequency, since the frequency stability of the klystron was insufficient to enable a musical note to be produced by the beating of two at nearly the same frequency. A wave-guide interferometer was therefore developed for this purpose, capable of detecting a change in frequency of 1 part in 6×10^6 . A full description of this instrument has already been given (Pippard 1949b).

Power from the oscillator was led through a short length of cable to a wave-guide, in which was a variable resistive strip attenuator of conventional design. This wave-guide was attached to the wave-guide which supplied power to the resonator. The wave-guide leaving the resonator was joined to a fixed attenuator and thence to the detector, which consisted of a standard crystal rectifier mounted in the wave-guide and tuned by means of a screw of variable position and depth of insertion into the guide. The rectified current was measured with a sensitive galvanometer.

PREPARATION OF SPECIMENS

In the preliminary measurements the specimens consisted of tin wires about 1.4 mm. long and ranging in diameter from 0.25 to 1 mm., which were prepared by casting in thin-walled glass tubes, the glass being subsequently removed. It is, however, difficult if not impossible to remove the glass without damaging the surface of the metal, and in these measurements it is of fundamental importance to employ only unstrained surfaces. This point has been clearly demonstrated by the work of Chambers (1950) on the anomalous skin effect in copper, in which he has shown that it is only possible to obtain results in accordance with theory by carefully removing all traces of surface strain, by annealing and electrolytic polishing. If the measurements of surface impedance are to be of any use they must be carried out on specimens for which the surface is smooth and the surface layers have the same crystalline structure as the interior. It is the neglect of this requirement which vitiates the results on drawn wires presented in II (see Pippard 1949*a*; Chambers 1950).

In the later measurements, therefore, careful attention was paid to the state of the surface. Some of the single-crystal specimens were made by casting pure tin in thin-walled silica tubes, which were not removed before performing the experiments, but which did not give rise to very serious dielectric losses in spite of their occupying regions of high field strength. Microscopic examination showed that the cast surfaces were very smooth, causing no appreciable scattering of light even under intense illumination. It is likely therefore that whatever irregularities existed were no larger than one-tenth of a wave-length, or 5×10^{-6} cm. There is still the possibility that the surfaces of the cast specimens were strained, especially if there is any tendency for the tin to stick to the walls of the tube, and for this reason a number of bare specimens were also prepared. Tin wires of about 0.5 mm. diameter were prepared by casting in glass, and removing the glass, and were then recrystallized in air, the crystal axis being oriented in the desired direction by means of a seed crystal. The oxide layer on the recrystallized wire was removed in hydrochloric acid, and the wire was then electrolytically polished in a solution consisting of 1 part of 20 % perchloric acid to 4 parts of ethyl alcohol. This polishing proved to be a troublesome process, but eventually a satisfactory technique was evolved, in which a small current (about $\frac{1}{2}$ A/cm.²) was allowed to flow for 10 sec., and was followed by a short pulse of a much larger current (40 A/cm.² for about $\frac{1}{10}$ sec.). After this treatment the wire was often found to present an extremely smooth appearance under microscopic examination, with no noticeable scattering of light, and the specimens were prepared from those sections which seemed smoothest. Not all the specimens were perfectly smooth, but none had rough patches which occupied more than a few per cent of the total surface, and care was taken to avoid rough patches at the centre of the specimen where the current density is highest. It is therefore believed that the results obtained with these specimens are results appropriate to an unstrained surface having no irregularities greater than 5×10^{-6} cm.

The polished specimens were verified to be single crystals by etching portions of the original recrystallized wire adjacent to each end of the specimen, and at the same time the orientation was determined. The cast specimens were verified to be single

crystals in the same way, but the orientation was finally determined by X-ray examination of the actual specimens.

EXPERIMENTAL PROCEDURE

The experimental procedure adopted was closely similar to that described in I, II and III, and the quantities determined were the normal surface resistance, R_n , just above the transition temperature (3.712°K on the 1932 temperature scale (Keesom 1932)), the ratio R/R_n for superconducting tin as a function of temperature between 3.712 and 1.7°K , and the difference in surface reactance of the superconducting and the normal metal (with a magnetic field applied to destroy superconductivity) as a function of temperature in the same range.

The normal resistance, R_n , above the transition temperature was determined by measuring the variation of Q with transmission coefficient, t , for the specimen in an unlined brass can, and extrapolating the linear relation so obtained to zero value of t . This gave Q_0 for the resonator. To allow for extraneous losses due to the can and to dielectrics the temperature was lowered as far as possible, when the specimen losses became negligible and the contribution of the extraneous losses could be determined accurately. As in the 25 cm. measurements it was not feasible to deduce R_n directly from Q_0 , on account of the difficulty of computing the conversion factor from the geometry of the resonator. A measurement of Q_0 was therefore made also at the temperature of solid carbon dioxide (-78.5°C), which is sufficiently high for the classical skin-effect theory to be valid. Since the skin resistance can here be calculated from the bulk resistivity the conversion factor is experimentally determined, and hence R_n at the transition temperature may be found.

For the ratio R/R_n of superconducting tin the procedure was exactly the same as described in I, the coupling being adjusted at each temperature to give rise to a standard value of t , usually 0.4 , so that relative values of Q were also relative values of $1/R$. A slight complication in these experiments arose from the heating of the specimen by the small amount of power which was dissipated in it, which although only a few microwatts, was sufficient to raise the temperature about 0.02°K . The analysis of the influence of this heating on the measurements is lengthy but straightforward, and will not be given here. It is sufficient to quote the result, that if the power dissipation be kept constant the only effect, apart from an easily ignorable bump at the transition temperature, is a bodily shift of the measured curve by an amount rather greater than the actual temperature rise of the specimen. A computed example is shown in figure 2, from which it may be seen that there is no appreciable distortion of the gradient of the curve. In all these measurements, then, a suitable amount was added to the temperature readings to bring the apparent transition temperature to 3.712°K , which appears from the measurements at 25 cm. (where the heating is unimportant owing to the greater bulk of the specimen) to be the transition temperature for this sample of tin.

The principle of reactance measurement, by observation of the change in resonant frequency resulting from destruction of superconductivity by means of a magnetic field, has been given in III, and the method adopted here was the same but for one

respect. In the 25 cm. measurements the factor for converting a given change in resonant frequency into a change of inductive skin depth (the inductive skin depth, δ_i , is defined as $X/4\pi\omega$, X being the surface reactance) was computed from the dimensions of the resonator. As is shown in the appendix, this conversion factor may be determined experimentally from observations at such a temperature that the classical skin-effect theory is valid. If a change of inductive skin depth from zero to δ_i occasions a change $\Delta\omega_i$ in resonant frequency, and a resistive skin depth δ_r corresponds to a resonance peak of width $\Delta\omega_r$ between half-power points, then these quantities for a given resonator of any shape are related by the equation

$$\frac{\delta_i}{\Delta\omega_i} = \frac{\delta_r}{\Delta\omega_r},$$

and each side of the equation is constant for a given resonator. At the temperature of solid carbon dioxide δ_r may be calculated from the known bulk conductivity, since $\delta_r = (2\pi\omega\sigma)^{-1/2}$ according to the classical theory of the skin effect, and $\Delta\omega_r$ may be measured in the usual way. Hence measured values of $\Delta\omega_i$ may be immediately converted into changes of inductive skin depth. It should be noted that the $\Delta\omega$'s need not be actual frequencies, but may be readings on the differential wavemeter, so that systematic errors in the wavemeter do not affect the accuracy of the experiment.

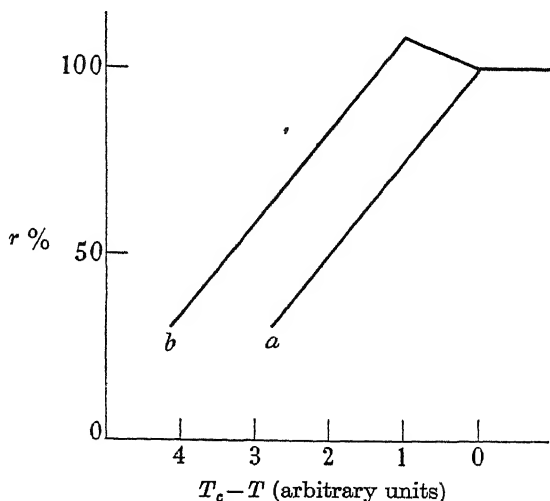


FIGURE 2. Effect of heating of specimen by measuring current. *a*, ideal variation of r with temperature; *b*, observed variation if heating at peak of resonance is 1 unit of temperature.

In this method of measuring surface reactance it is of prime importance that the specimen should exhibit the Meissner effect, so that when the magnetic field is removed there shall remain trapped in the specimen no flux which may keep regions of the surface in the normal state. This matter was studied with several of the specimens in separate experiments. The specimen was cooled below the transition temperature in the absence of a field; the resistance was then measured, and again after superconductivity had been destroyed and recreated by switching a magnetic field on and off. In no such experiment was any change observed, and it was con-

cluded that if any part of the surface remained normal it amounted to less than $\frac{1}{2}\%$ and would result in no appreciable error.

EXPERIMENTAL RESULTS

(a) Resistance of normal tin

Ten specimens were used, of which five were cast in silica tubes and five were electrolytically polished. In figure 3 the surface conductivity Σ ($= 1/R$) just above the transition temperature is displayed as a function of θ , the angle between the tetrad axis of the crystal and the axis of the specimen, which is also the direction of current flow. For tin of the purity of the samples used in these experiments the mean free path of the conduction electrons is very many times as great as the classical skin depth, the former being about 5×10^{-3} cm. and the latter about 2×10^{-6} cm. Under these conditions, referred to by Reuter & Sondheimer (1948) as 'extreme anomalous conditions', Σ has very nearly reached its limiting value and is almost independent of the bulk conductivity. The departure from the smooth curve of one point (at 4°) cannot therefore be explained as arising from a slightly different value of the bulk conductivity. It is more likely that this specimen, which was recrystallized after casting in its silica tube and which showed surface faults, was much rougher microscopically than the others, and it is to this fact that its lower conductivity is to be attributed. Apart from this point there is good agreement between the results obtained with specimens prepared in the two different ways, and this suggests that the problems of surface roughness and strain have been successfully overcome.

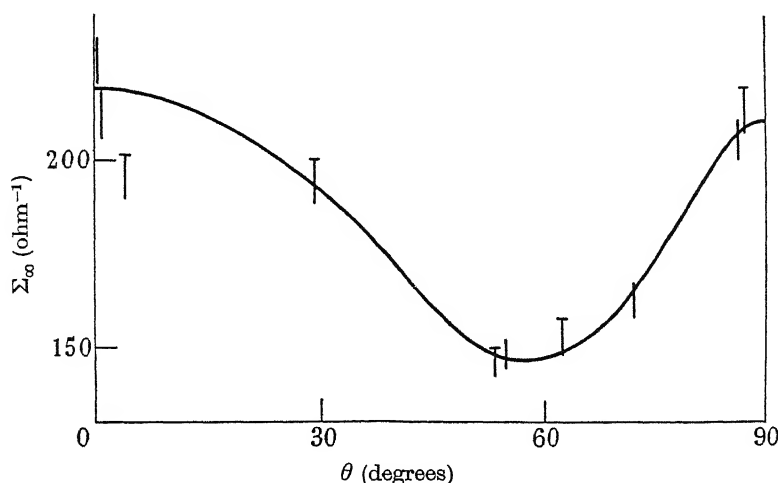


FIGURE 3. Variation with orientation of surface conductivity of normal tin. The plain lines refer to electrolytically polished specimens, the rest to cast specimens.

For an isotropic metal in which collision of an electron with the surface terminates a free path ($p = 0$ in the notation of II), Reuter & Sondheimer deduce for the limiting value of Σ , $\Sigma_\infty = 0.568 \left(\frac{ne^2}{m\nu\omega^2} \right)^\dagger$, in which n and m are the number per unit

volume and effective mass of the conduction electrons, and v is the Fermi velocity. The conductivity should therefore vary as $\omega^{-\frac{2}{3}}$ for a given material under extreme anomalous conditions. Since the anisotropy of the effect was not studied at 25 cm. it is not possible to make a detailed comparison of the results at the two frequencies, but if it be assumed that the specimen used for the 25 cm. measurement consisted of a random polycrystalline aggregate the value of 678 ohm^{-1} quoted in II may be compared with an average of the results exhibited in figure 3, due account being taken of the statistical weight of each orientation. This average value is 174 ohm^{-1} , which agrees very well with the value 172 ohm^{-1} calculated from the 25 cm. measurements. It is of interest to note that Fairbank (1949), using cavity resonators at 3.2 cm., obtained a rather lower value of the surface conductivity, about 115 ohm^{-1} , and that the value obtained by Maxwell *et al.* (1949) at 1.25 cm. was also lower than would be expected from the present experiments. It is likely that the explanation of these discrepancies is to be found in the method of preparation of the specimens used by these workers; the lattice distortions produced by milling and hobbing will be similar to those produced by drawing, which are known from the work of Chambers to be responsible for serious errors, in the noble metals at any rate.

The form of the anisotropy of Σ_∞ is interesting, in that it differs from the result to be expected for a classical metal. In an anisotropic classical metal the usual conductivity equation, $\rho \mathbf{J} = \mathbf{E}$, must be generalized by treating ρ as a symmetric second-order tensor. For a uniaxial crystal such as tin the resistivity tensor may be represented by an ellipsoid of revolution, having axial symmetry about the tetrad axis. The exact treatment of the classical skin effect in such a crystal is tedious, but not difficult, and only the result will be quoted. For a cylindrical specimen such as those used in this work the measured surface resistance will vary with orientation in a manner which is represented to a close approximation by the equation

$$R^2 = 2\pi\omega\{\rho_1 + (\rho_2 - \rho_1) \sin^2 \theta\},$$

in which ρ_1 and ρ_2 are the principal resistivities parallel and perpendicular to the tetrad axis. For tin at room temperature this equation is in error by something less than 1 %. Consequently if the metal behaved according to the classical theory of the skin effect, R and consequently Σ should be monotonic functions of θ , which is clearly not so in the present experiments.

The reason for this behaviour is to be found in the form of the equations connecting \mathbf{J} and \mathbf{E} in the classical and anomalous theories. In the former the relation is, for an isotropic metal, a linear point-relation, by which the value of \mathbf{J} at any point is determined solely by the value of \mathbf{E} at that point, and in an anisotropic metal this character is retained when a resistivity tensor is introduced. In the anomalous theory, on the other hand, the relation between \mathbf{J} and \mathbf{E} , while still linear, is no longer a point-relation, since \mathbf{J} is determined by the configuration of the electric field over the whole metal (see Reuter & Sondheimer 1948, equation (14)). There is now no longer any quantity analogous to ρ which may be simply generalized by the introduction of a tensor, and as a consequence the mathematical formulation of the problem permits a much greater freedom of behaviour in an anisotropic material than is allowed by a point-relation.

A rough physical picture of the behaviour may be obtained by means of the ineffectiveness concept, introduced in II and justified to some extent by Reuter & Sondheimer. When the mean free path of the electrons is very much greater than the skin depth the greater part of the current is carried by that small fraction of the electrons moving at such glancing angles to the surface that they may execute a substantial part of a free path while remaining within the skin depth. It is the density and effective mass of these electrons for a given crystal orientation which determine the surface resistance, and since in a metal such as tin the shape of the Fermi surface is likely to be very different from the spherical shape assumed in the free-electron theory of metals, it is to be expected that for different orientations the surface resistance may vary considerably, and in no simple manner. From this point of view it is easy to see that even with the geometry of the present experiment, in which the direction of current flow is the same at all points of the specimen, the results obtained represent some sort of average effect, since at different azimuths of the specimen different groups of electrons will be effective in determining the resistance. A complete picture of the anisotropic effects can only be obtained from experiments on plane surfaces, and it is not possible to draw any detailed conclusions concerning the shape of the Fermi surface from the results of figure 3. It seems likely, however, from the maxima of Σ_{∞} at the extreme orientations, either that the directions of motion of the electrons are concentrated around the crystal axis and the plane normal to the axis, or that in those directions the effective mass is much smaller than in intermediate directions.

(b) *Reactance of superconducting tin*

Reactance measurements were carried out on the same ten specimens which were used for the measurements just described on normal tin. The difference in inductive skin depth between the superconducting and normal metals was determined as a function of temperature, and since for tin the inductive skin depth of the normal metal is independent of temperature in this range any variation of the observed difference with temperature must be ascribed to the superconducting metal. Two typical curves are shown in figure 4, in which the difference in inductive skin depths is plotted against the reduced temperature, $t (= T/T_c$, where T_c is the transition temperature). The skin depth in the normal metal is designated δ_n , in the superconductor λ' ; from the definition of inductive skin depth it follows that at low frequencies λ' is equal to the penetration depth, λ . It will be seen from these curves that the method is capable of considerable accuracy, and in fact under the best conditions the measured probable error of a single determination of $\delta_i - \lambda'$ was no greater than 4×10^{-8} cm., about 1 % of the penetration depth at 0° K.

At the transition temperature $\delta_i - \lambda'$ vanishes, as is to be expected, and with lowering of the temperature the difference increases towards a limiting value, $\delta_i - \lambda'_0$, which is different for each orientation. The relation between these curves and the variation with temperature of the superconducting penetration depth has been discussed in III, and will be further discussed in the next paper. Here we shall be concerned only with presenting the experimental results, a task which is greatly simplified by the observation that for values of t up to 0.90 the curves of $\lambda' - \lambda'_0$ for

all specimens have the same shape, apart from a scaling factor characteristic of each orientation. This is illustrated in figure 5, where $\lambda' - \lambda'_0$ has been plotted against an empirically determined function of temperature, $\phi(t)$, to which a close analytical approximation is given by the formula

$$\phi(t) \doteq 14y - 5y^2, \quad \text{where} \quad y = (1 - t^4)^{-\frac{1}{2}} - 1. \quad (1)$$

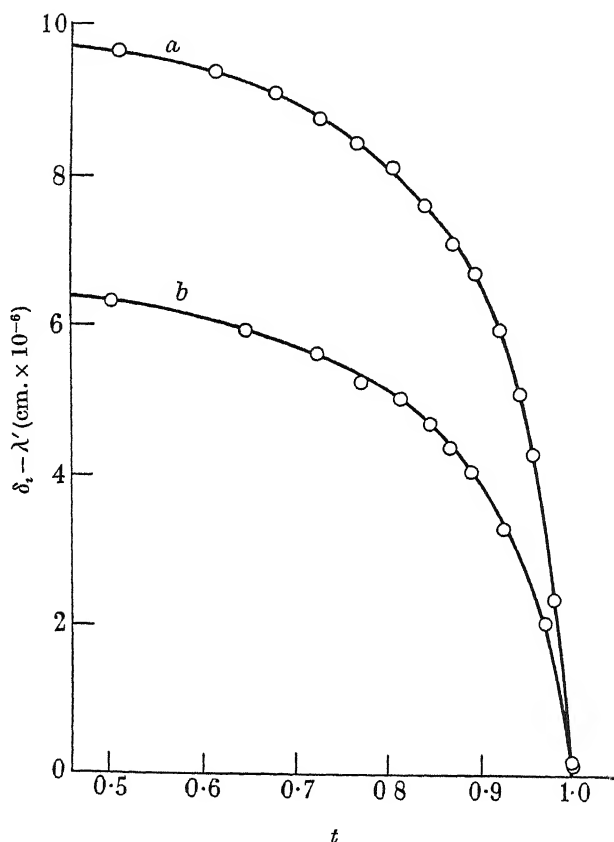


FIGURE 4. Temperature variation of inductive skin depth. *a*, $\theta = 62.5^\circ$; *b*, $\theta = 0^\circ$.

Plotted in this way, the points for each specimen lie on a straight line, and we may take the gradient, g , of the line to characterize the variation of λ' with temperature for a given specimen. The results of the reactance measurements are thus summarized in figure 6, which shows the variation of g with θ . The variation of λ' with temperature is slowest for the extreme orientations, and most rapid around the value 60° of θ ; this behaviour is similar to the variation with θ of the resistive skin depth in the normal metal, which is inversely proportional to Σ as exhibited in figure 3. The lines in figure 6 represent estimated probable errors from all causes, of which errors due to experimental scatter of the points in figure 5 and errors in the determination of the conversion factor between frequency change and inductive skin depth are the most important.

Now from equation (1) it is clear that at very low temperatures λ' is a linear function of y , in agreement with the conclusions of Daunt, Miller, Pippard & Shoenberg (1948) that the available evidence favours a law for the variation of λ with t of the form

$$\lambda = \lambda_0(1 - t^4)^{-\frac{1}{2}}. \quad (2)$$

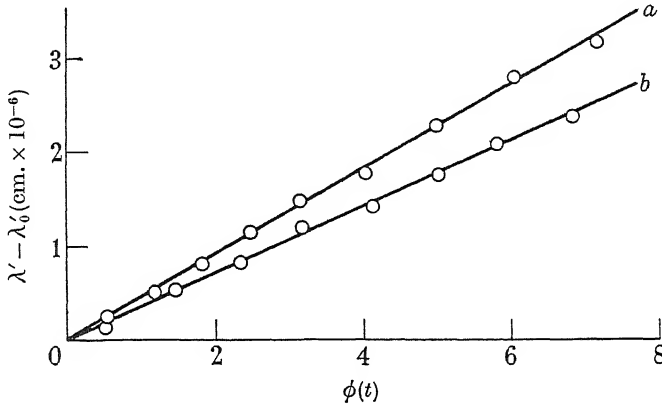


FIGURE 5. Linear plot of temperature variation of inductive skin depth.

$a, \theta = 62.5^\circ$; $b, \theta = 0^\circ$.

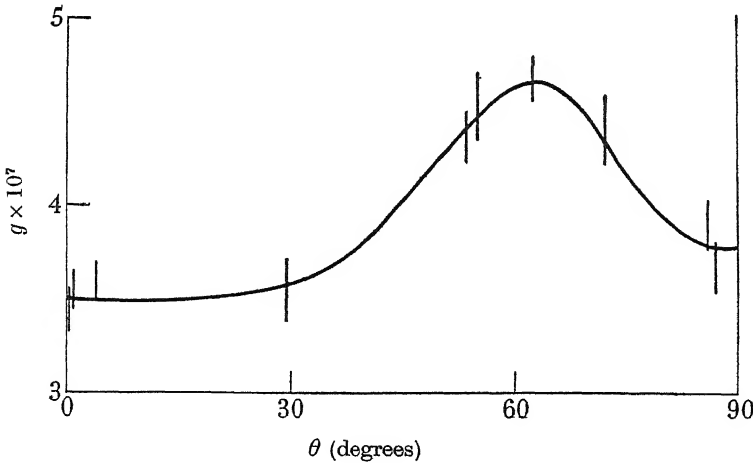


FIGURE 6. Anisotropy of inductive skin depth.

In the present experiments λ' cannot be the same as λ right up to the transition temperature, since λ tends to infinity while λ' tends to δ_i . It is therefore of interest to see whether the quadratic term in $\phi(t)$ represents a departure of the penetration depth from the simple law, hitherto unobserved through experimental inaccuracies, or whether it is due simply to the same cause as limits λ' to values not greater than δ_i , that is, the mechanism which is responsible for the high-frequency resistance. To test this point careful reactance measurements were made on a tin specimen at 25 cm., and the result of this experiment is shown schematically in figure 7. Here values of $\lambda' - \lambda_0'$ at three different frequencies are plotted against $(1 - t^4)^{-\frac{1}{2}}$, account being

taken of the differences in the specimens by scaling the measured values so as to bring the initial gradients into coincidence. It is clear that the departures from the law represented by equation (2) are to be ascribed entirely to the effects of the high frequency; the interpretation of this result will be discussed quantitatively in part V.

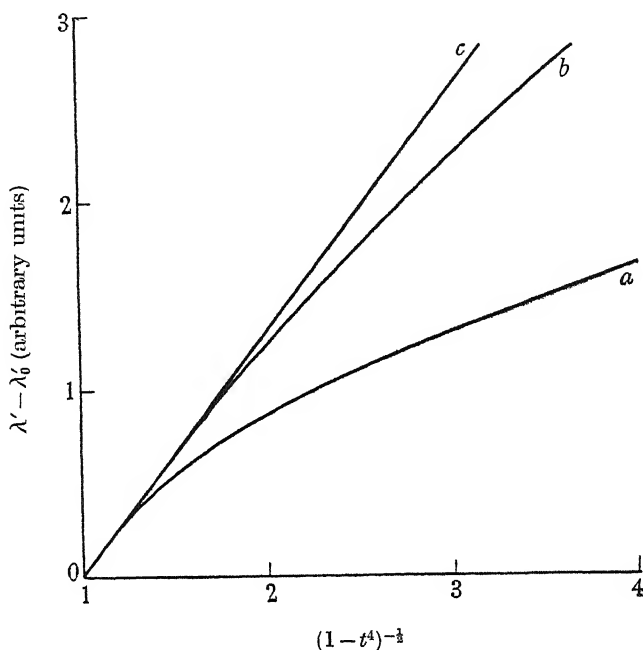


FIGURE 7. Temperature variation of inductive skin depth at different frequencies. *a*, 9400 Mc./sec.; *b*, 1200 Mc./sec.; *c*, 70 c./sec. (Laurmann & Shoenberg 1949).

The theory of the anomalous skin effect predicts that in a normal metal under extreme anomalous conditions the surface reactance and resistance should be in the ratio $\sqrt{3}$, rather than unity as in the classical theory, and these measurements enable the ratio to be estimated. Experimentally what may be determined is the difference between δ_i and λ'_0 , and if it be assumed that at the lower temperatures, where the curves of figure 8 become linear, λ' is equal to λ , it is possible to calculate λ_0 from the data. In fact, inspection of equations (1) and (2) shows that $\lambda_0 = 14g$. Hence δ_i may be calculated, and in figure 8 is shown the variation with θ of δ_i , together with the variation of δ_r calculated from the data shown in figure 3. In the same diagram the ratio δ_i/δ_r , deduced from the smoothed curves, is shown. Since δ_i is defined as $X/4\pi\omega$ and δ_r as $R/2\pi\omega$, the ratio δ_i/δ_r should be $\frac{1}{2}\sqrt{3}$, or 0.866, according to the theory of Reuter & Sondheimer. On account of the scatter of the experimental points the exact shape of curve is doubtful, although the general form is probably reasonably correct. Since the theory of the anomalous skin effect has not been extended to an anisotropic metal it is not possible to say whether this variation of δ_i/δ_r is to be expected. However, the mean value of δ_i/δ_r , appropriate to a polycrystalline specimen, calculated from this curve, is 0.88, in quite as good agreement with the theoretical

value as may be expected, so that these results may be said to confirm the theory of Reuter & Sondheimer.

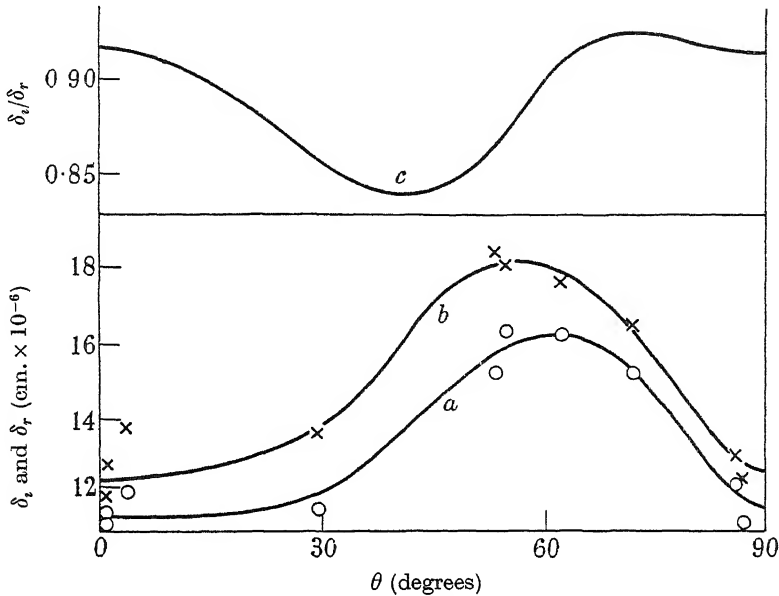


FIGURE 8. Anisotropy of inductive and resistive skin depths in normal tin.
 a, the inductive skin depth, δ_i ; b, the resistive skin depth, δ_r ; c, δ_i/δ_r .

(c) Resistance of superconducting tin

The variation with temperature of the surface resistance of superconducting tin is similar at 3.2 cm. to that at 25 cm. described in I. As the temperature is lowered through the critical temperature the resistance begins to fall sharply, and continues to fall steadily, though at a slower rate, with further lowering of the temperature. As would be expected, the rate of decrease of resistance is less at the higher frequency, and it is consequently easier to obtain accurate data. The variation with temperature of r , the ratio of the resistance R at a temperature t to the resistance R_n just above the transition temperature, was determined for seven of the monocrystalline specimens, and the results for two specimens are shown in figure 9, together with the curve obtained previously at 25 cm. These results have been corrected for the residual cavity losses in the following way. In the preliminary experiments it was found that two specimens whose diameters differed by a factor 4 gave curves of almost precisely the same shape, except that r tended to different values at 0°K. From a separate experiment in which the specimen was a lead-coated brass wire an estimate was made of the losses due to the can, and these losses were discovered to be much greater than any losses due to the dielectric supports, which could be neglected. It was therefore possible to predict the ratio of the residual cavity losses for the two specimens, and this ratio turned out to be the same as the ratio of the limiting values of r . This result makes it clear that the greater part of the limiting value of r is due to the can and not to the specimen. Further confirmation was obtained by an experiment on a very fine tin wire, of diameter 50 μ , cast in a thin-

walled silica tube. The residual value of r for this specimen was only 0.22 %, of which 0.10 % was calculated to be the contribution of the can losses. Thus the value of r for the specimen must certainly fall at least to 0.12 %, and it requires a power factor of only 10^{-5} for the silica tube to account for this residual loss. We may be sure, therefore, that r falls below 0.1 % and probably vanishes at the absolute zero. For all the specimens measured the limiting value of r could be accounted for as being due to can and dielectric losses, and the measured values have been corrected on this assumption so as to make each curve tend to zero at 0° K.

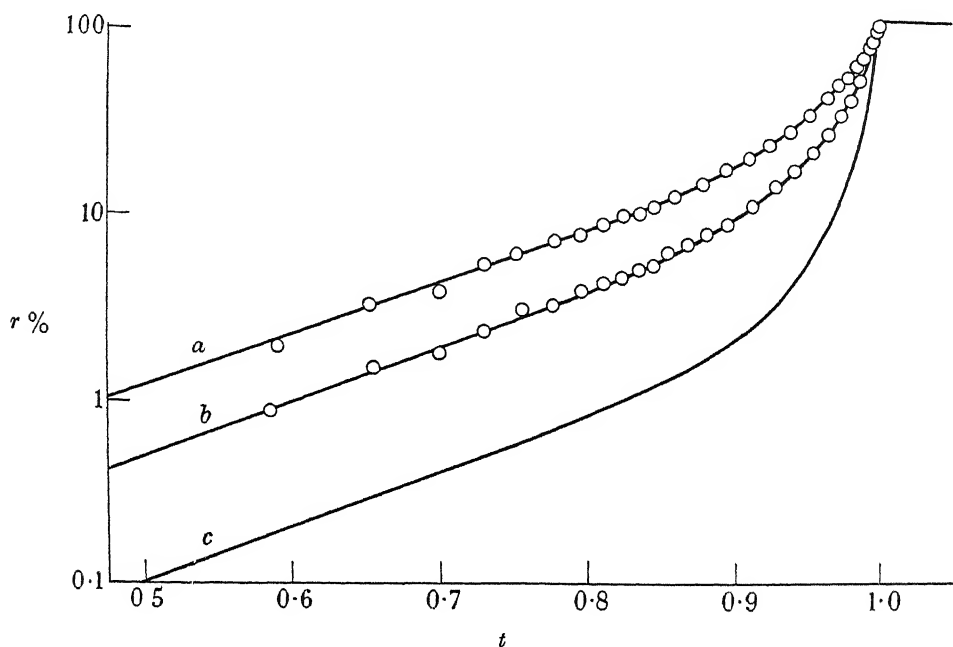


FIGURE 9. Temperature variation of superconducting resistance. a , 3.2 cm., $\theta = 87^\circ$; b , 3.2 cm., $\theta = 62.5^\circ$; c , 25 cm., average of several specimens.

This is a different result from that of the 25 cm. measurements, where it appeared that r tended to a value 0.7 % after due allowance had been made for the cavity losses. It is most improbable that if r tends to zero at 3.2 cm. it tends to a non-zero value at 25 cm., and it is therefore concluded that the methods of estimation of the correction in the earlier work were somewhat in error. A new measurement on a fine specimen at 25 cm. has made it clear that the limiting value of r is certainly not greater than 0.4 %, but in order to ascribe this remaining resistance to the silica sheath of the specimen it is necessary to assume a power factor at 25 cm. of 5×10^{-5} , five times greater than at 3.2 cm. No experiments have been performed to discover whether this is the true explanation, but for the present purpose it has been assumed that r tends to zero at both frequencies, and the results presented in I have been corrected accordingly.

As with the reactance measurements, it has been found that the resistance curves for different orientations show certain similarities which facilitate the presentation of the experimental results. In the region just below the transition temperature, in

which r varies from 100 to 40 %, all the curves have the same shape, apart from a scaling factor applied along the temperature axis, so that the results may be expressed in the form

$$\Delta T = \text{constant} + G_1 \psi(r), \quad (3)$$

in which ΔT is the difference between the transition temperature and the temperature measured, G_1 is a factor characteristic of each specimen, and $\psi(r)$ is an empirically determined function whose form is very nearly represented by the equation

$$\psi(r) \doteq 800 - 435 \log_{10}(100r - 32). \quad (4)$$

Examples of this method of treating the results are given in figure 10.

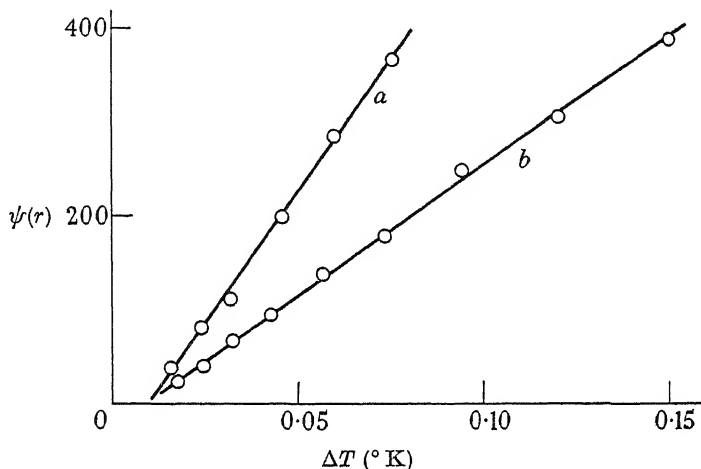


FIGURE 10. Linear plot of resistance just below the transition temperature.
 a , $\theta = 55^\circ$; b , $\theta = 87^\circ$.

Similarly, at the lower temperatures, when r is less than about 7 %, all the specimens give curves of the same shape apart from a scaling factor, applied this time along the resistance axis. Thus r may be expressed in the form

$$r = G_2 f(t), \quad (5)$$

in which G_2 is a factor characteristic of each specimen, and

$$f(t) = \frac{t^4(1-t^2)}{(1-t^4)^2}. \quad (6)$$

It is noteworthy that the results at both frequencies can be expressed in this way, as is shown in figure 11. The extreme curve at 3.2 cm. (for $\theta = 87^\circ$) is seen to deviate from linearity when r exceeds 7 %, but otherwise all the specimens give straight lines for values of $f(t)$ up to 0.7, corresponding to a value 0.86 of t .

In this way the shape of the resistance curve may be described over the greater part of its length in terms of two parameters, G_1 and G_2 , and it is not surprising to find a close correlation between these two parameters. In figure 12, G_2 is plotted against G_1 , and it may be seen that within the limits of experimental error the

parameters are proportional to one another. This result will play an important part in the analysis of the measurements to be given in part V. On the same curve is shown a point obtained at 25 cm. It is difficult to say how accurate this point is, especially as regards the value of G_1 , which may be somewhat larger than its true

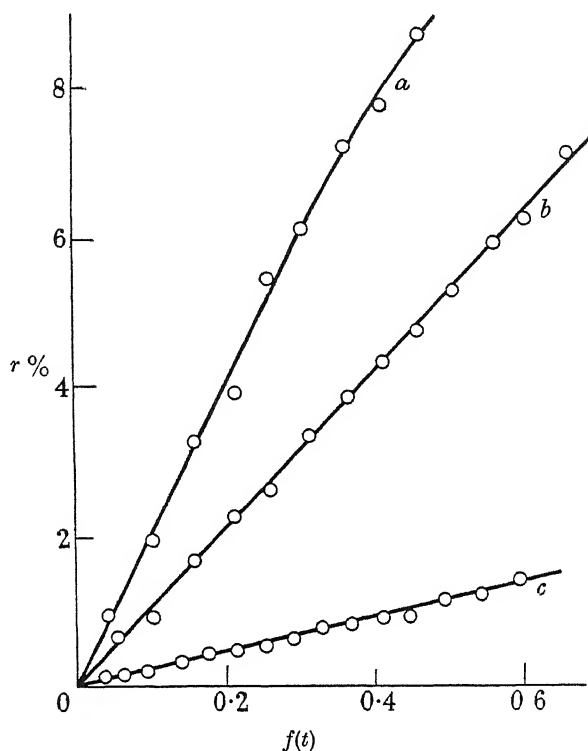


FIGURE 11. Linear plot of low-temperature resistance. *a*, 3.2 cm., $\theta = 87^\circ$; *b*, 3.2 cm., $\theta = 72^\circ$; *c*, 25 cm., orientation unknown.

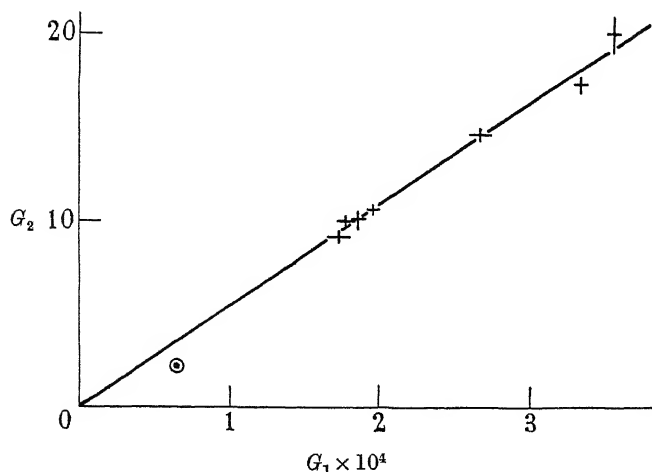


FIGURE 12. Correlation of G_1 and G_2 . The circle refers to a measurement at 25 cm., the crosses to measurements at 3.2 cm.

value on account of the inherent breadth of the transition, even for stationary fields. It is unlikely that the true position of the point is farther away from the straight line than the point shown in the diagram, and more than likely that it is much closer. Finally, the anisotropy of the superconducting resistance may be exhibited by plotting either G_1 or G_2 against θ , and in figure 13 the angular variation of G_2 is shown. The characteristic humped curve appears again, this time in opposition to the resistance curve of the normal metal; for those orientations around 60° for which the normal resistance is highest, the resistance ratio r falls most steeply below the transition temperature, so that eventually it is these orientations which show the smallest resistance.

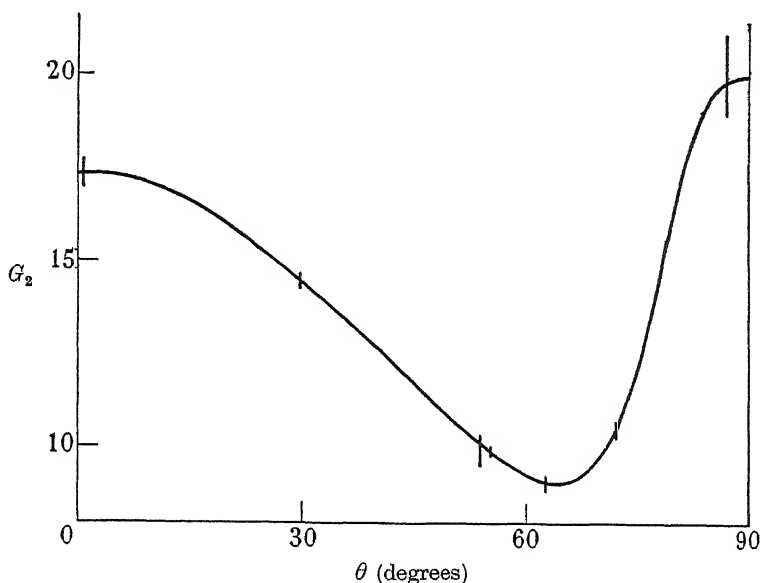


FIGURE 13. Anisotropy of superconducting resistance.

DISCUSSION

The detailed discussion of the results presented here and in the earlier parts of this series is a rather lengthy matter, which will be undertaken in part V. For the moment we shall concern ourselves only with a discussion of the justification for regarding the results as truly representative of the behaviour of ideal specimens. Every effort has been made in the preparation of the specimens to ensure that the surfaces shall be smooth and strain-free, but if these efforts have not been successful it is clear that quantitative deductions from the results will have little or no value. As regards the possibility of strains being present in the surface layers, it seems likely that the methods of preparation are such as to eliminate these. The work of Chambers (1950) has shown that it is possible to achieve good agreement between experiment and theory in measurements of the anomalous skin effect in copper, a metal which is much less readily annealed than tin, by electrolytically removing the surface layers which cause serious errors in drawn wires. Since in the present work the specimens were recrystallized from the melt before electrolytic polishing there is even less possibility

of the measured surfaces being sufficiently strained to affect the results. Additional evidence in favour of this conclusion is provided by the agreement between experimental and theoretical values of δ_s/δ_r , and by the fact that the resistance in the superconducting state appears to tend to zero at 0° K. From the work of Maxwell *et al.* (1949) and Fairbank (1949) it seems that the residual resistance of the superconductor is a sensitive test of surface strain.

It is not quite so easy to dispose of the possibility that the surfaces are rough to an extent which would be important in this work, since the thickness of the layer in which the current flows is only about 10^{-5} cm. above the transition temperature, and falls to half this value in the superconducting state. It is to be expected, then, that roughness on this scale will affect the results, and it might be regarded as significant that those orientations which show the largest normal resistance and superconducting inductive skin depth are just those for which the orientation of the crystal with respect to the surface is such as to encourage microscopic roughness. However, there are several arguments against this hypothesis. First, there is good agreement between the results obtained with cast and with polished surfaces, and it is improbable that the same degree of roughness would result from both methods of preparation. Secondly, in order to account for the shape of the curve of figure 3 the scale of the roughness would have to be several times as great as the skin depth, according to the calculations of Morgan (1949), and thus comparable to the wave-length of visible light and easily detectable by the scattering it would produce. Thirdly, if the scale of the roughness is rather less than 10^{-5} cm it would have little effect on the normal resistance, but might be sufficient to increase the apparent superconducting resistance and reactance, thus accounting for the bump in the curve of figure 6. However, if this were so we should expect the superconducting resistance to show a corresponding increase at orientations around 60°, while in fact these are the orientations which give the lowest values of resistance.

To sum up, we may conclude that while there is still a possibility that surface roughness may play a small part in determining the measured values of surface impedance, there is strong evidence against its being the cause of the observed anisotropic behaviour, and there is therefore some justification for proceeding to an analysis of the results on the assumption that they represent the ideal behaviour of monocrystalline tin.

APPENDIX

Consider any system of conductors forming a resonator, and let us suppose first that all the conductors are perfect, so that there is no field penetration, and no decrement. Under these conditions there will be a certain distribution of surface-current density on the conductors, which we shall write as $j(s)$. At one particular phase of the vibration the current, and hence the magnetic field, will at all points reach a maximum value, and at the same time the charges on the surfaces of the conductors will vanish, together with the electric field. The energy of the resonator will then be wholly magnetic, and may be written in the form

$$U_{\text{mag.}} = \frac{1}{2} L_0 \iint_S j^2 ds,$$

the integration being over all the conductors. This equation formally defines the self-inductance, L_0 , of the resonator.

One-quarter period after this current maximum, the currents and magnetic field are everywhere zero, and the energy is stored in the electric field arising from charge distributions on the conductors. Writing the surface charge density as $q(s)$, we may relate q and \mathbf{j} by the continuity equation

$$\dot{q} = -\text{div}_s \mathbf{j},$$

where div_s is written for the surface divergence; or, for a vibration of frequency ω ,

$$q = \frac{i}{\omega} \text{div}_s \mathbf{j}.$$

Thus the form of the charge distribution is determined by \mathbf{j} and its magnitude varies inversely as ω . Consequently we may write for the electrical energy, which is proportional to the square of the charge density,

$$U_{\text{el.}} = \frac{1}{2C_0\omega^2} \iint_S j^2 ds,$$

and so define the self-capacity, C_0 , of the resonator. The natural frequency of oscillation is obtained by equating $U_{\text{mag.}}$ and $U_{\text{el.}}$ to give

$$\omega_0^2 = 1/L_0C_0.$$

Now let us suppose that the magnetic field penetrates into the conductors, while at the same time the current distribution remains unaltered, this latter assumption will be discussed later. There will result an extra term $\Delta U_{\text{mag.}}$ in the magnetic energy, while the electrical energy will remain unchanged, since the charge will still be carried on the surface. The resonant frequency will be obtained by equating $U_{\text{el.}}$ to $U_{\text{mag}} + \Delta U_{\text{mag.}}$, so that the change in resonant frequency will be given by the equation

$$\frac{\Delta\omega_i}{\omega_0} = -\frac{\Delta U_{\text{mag.}}}{2U_{\text{mag.}}}.$$

To evaluate $\Delta U_{\text{mag.}}$, we consider the definition of inductance in terms of stored magnetic energy, and obtain immediately

$$\Delta U_{\text{mag.}} = \frac{1}{2} \iint_S L j^2 ds,$$

where L is the surface inductance. If the material of the resonator is uniform, L is constant, and we obtain the result

$$\frac{\Delta\omega_i}{\omega_0} = -\frac{L}{2L_0}.$$

Now consider the decrement of the resonator, for which the surface resistance will be taken as R . The energy dissipated per cycle is readily seen to be given by

$$\dot{U}_w = \frac{\pi R}{\omega} \iint_S j^2 ds,$$

and hence from the definition of Q_0 ,

$$Q_0 = \frac{2\pi U_{\text{mag.}}}{\dot{U}_\omega} = \frac{\omega_0 L_0}{R}.$$

If a frequency change $\Delta\omega_r$ traverses the resonance peak between the half-power points, then Q_0 is also given by $\omega_0/\Delta\omega_r$. Hence

$$\left| \frac{\Delta\omega_i}{\Delta\omega_r} \right| = \frac{\omega_0 L}{2L_0 R} = \frac{\omega_0 L}{2R}.$$

The resistive skin depth δ_r is defined as $R/2\pi\omega_0$, and the inductive skin depth δ_i as $L/4\pi$. Hence $\frac{\Delta\omega_i}{\Delta\omega_r} = \frac{\delta_i}{\delta_r}$, which is equivalent to the expression quoted in the text. It should be noted that it has been assumed that the resistive losses in the resonator are entirely due to those conductors for which δ_i is considered. This means in practice that the value of $\Delta\omega_r$ employed must be the measured value corrected for residual losses in the can and dielectrics, in a way described in the text.

Finally, we must return to the assumption that no redistribution of current occurs when the surface inductance is changed. The justification of this assumption is that it is of only secondary importance whether or not small redistributions occur, since according to Rayleigh's principle this method of determining the resonant frequency always gives a value which is higher than the correct value if an incorrect mode of oscillation is considered. In particular, the frequency takes a stationary value with respect to small variations about the correct mode, and since we have considered the real mode of oscillation for the particular case of $\delta_i = 0$, any small changes incurred by the introduction of a non-zero value of δ_i will have only a second-order effect on the frequency, which may safely be ignored provided that $\Delta\omega_i$ is much smaller than ω_0 .

In carrying out the experimental work described above I received much help from Mr R. G. Chambers, who also kindly put his apparatus at my disposal for the new measurements at 25 cm., and Mr J. V. Smith and Mr E. Laurmann gave valuable assistance in the determination of the orientation of the specimens. My thanks are due to the Chief Superintendent, R. R. D. E. Malvern, for the loan of apparatus, and to Mr F. Sadler for help in the design and construction of apparatus and for the maintenance of an adequate supply of liquid helium.

REFERENCES

- Chambers, R. G. 1950 *Nature*, **165**, 239.
 Daunt, J. G., Miller, A. R., Pippard, A. B. & Shoenberg, D. 1948 *Phys. Rev.* **74**, 842
 Fairbank, W. M. 1949 *Phys. Rev.* **76**, 1106.
 Keesom, W. H. 1932 *Commun. Phys. Lab. Univ. Leiden*, Suppl. **71 d**.
 Laurmann, E. & Shoenberg, D. 1949 *Proc. Roy. Soc. A*, **198**, 560.
 Maxwell, E., Marcus, P. M. & Slater, J. C. 1949 *Phys. Rev.* **76**, 1332.
 Morgan, S. P. 1949 *J. Appl. Phys.* **20**, 352.
 Pippard, A. B. 1947 *Proc. Roy. Soc. A*, **191**, 370, 385 and 399 (parts I, II and III).
 Pippard, A. B. 1948 *Nature*, **162**, 68.
 Pippard, A. B. 1949a *Physica*, **15**, 40.
 Pippard, A. B. 1949b *J. Sci. Instrum.* **26**, 296.
 Reuter, G. E. H. & Sondheimer, E. H. 1948 *Proc. Roy. Soc. A*, **195**, 336.

Liquid helium films

I. The thickness of the film

By K. R. ATKINS, *The Royal Society Mond Laboratory, Cambridge*

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 29 March 1950)

The thickness (d) of the helium II film and its variation with height (H) and temperature were measured by a dynamic method involving the oscillations of a meniscus in a capillary. The variation with height could be represented only approximately by the equation $d = k/H^n$, as the effective value of n was greater at smaller values of H . The mean value of n over a range of heights from 0.5 to 5 cm. was 0.14, which is appreciably smaller than the values predicted by the theories so far advanced to explain the formation of the film. The order of magnitude of k was 2×10^{-6} cm., but it varied slightly with the nature of the surface or some other experimental factor.

1. INTRODUCTION

The theories of Tisza (1938, 1947) and Landau (1941) suggest that the transport properties of liquid helium II can be explained in terms of the two component hydrodynamics of a mixture of normal and superfluid liquids. The aim of the research described below was to throw further light on the exact nature of this hydrodynamics by studying the superfluid flow of the liquid helium film. The present paper describes how the thickness of the film and its variation with height were measured by a new method involving oscillations of the film, while part II will deal with the flow properties of the film and in particular the way in which its rate of flow was influenced by variations in its thickness. The problem of the thickness of the film also presents certain novel features of intrinsic interest associated with the equilibrium of a thick film in contact with its bulk liquid.

Previous measurements of the film thickness should be considered in the light of the recent discovery of Bowers & Mendelssohn (1949) that film phenomena are very sensitive to small quantities of solidified gas on the surface holding the film, and it is by no means clear in which of the measurements these contamination difficulties were avoided. In the original experiments of Daunt & Mendelssohn (1939) the film was evaporated off a large surface and collected in a capillary. They found no evidence for a film thicker than 10^{-7} cm. above the λ -point, and obtained a mean value of 3.6×10^{-6} cm. below the λ -point, though their individual observations varied erratically between 2.7×10^{-6} and 4.9×10^{-6} cm. These figures refer to a film extending from 3 to 18 cm. above the surface of the bulk liquid.† Kikoin & Lasarew (1939) used a similar method and obtained a result of the same order of magnitude, but no details of their method are available. More recently, Burge & Jackson (1949) have investigated the variation of film thickness with height and temperature using an optical method. They found that over most of the temperature range between 1.1° K and the λ -point the thickness was approximately 1.9×10^{-6} cm., but in a small

† I am indebted to Dr K. Mendelssohn for this information.

range of temperatures near the λ -point it decreased rapidly and they could find no evidence for the existence of the film above the λ -point. Assuming the thickness to vary inversely as the n th power of the height, they obtained values of n varying from 0.28 at the lowest temperatures to 0.4 near the λ -point. All these figures are provisional until a more accurate calibration of the optical constants of their apparatus can be made. Kistemaker (1947) approached the problem from a different direction by studying the adsorption of helium gas on glass at pressures near the saturated vapour pressure. He was able to produce a film as thick as 5×10^{-7} cm. above the λ -point, and as the temperature was lowered below the λ -point the thickness steadily increased but did not rise above 10^{-6} cm. There was no measurable change in film thickness as he raised the pressure from 95 to 99 % of the saturated vapour pressure. Long & Meyer (1949), however, have found that for the case of adsorption on jeweller's rouge the amount of gas adsorbed increases rapidly as the saturated vapour pressure is approached, and they advance reasons for believing that this is not due to capillary condensation. It is clear that there are many inconsistencies between these various experiments.

There are three theories of the formation of the film, all of which assume static equilibrium between the film and the bulk liquid. Schiff (1941) and, independently, Frenkel (1940) assume that the film is due entirely to the van der Waals forces of attraction between the helium atoms and the wall, and deduce that the film thickness should vary inversely as the cube root of the height. Bijl, de Boer & Michels (1941) consider the dominant factor to be the zero-point energy of the atoms confined within the small dimensions of the film, and this implies a film thickness varying inversely as the square root of the height. Temperley (1949) has attempted a rigorous treatment combining the van der Waals forces and quantum effects, but unfortunately his results are very sensitive to the exact values of certain parameters and he can deduce only the order of magnitude of the variation of thickness with height.

2. THEORY OF THE OSCILLATION METHOD

(a) *The principle of the method*

Consider, as in figure 1, a vessel containing liquid helium II emptying through the film. The inner level eventually comes to rest at the same height as the outer level, but, as the rate of flow is quite large even for very small level differences, this equilibrium position is approached with considerable momentum so that the inner level overshoots and then oscillates about its equilibrium position. This oscillation can be considered as a periodic interchange of energy between the kinetic energy of the film and the potential energy of the liquid inside the vessel which is raised above the outer level, and the period depends primarily on the thickness of the film. This type of oscillation was first discovered by Allen & Misener (1939), not for film flow, but for flow through narrow subsurface channels, and was used by them to estimate the cross-sectional area of the channels. This estimate agreed to within 30 % with an independent estimate involving the rate of flow of helium gas through the channels, and so the method can be relied upon to give at least an order of magnitude for the thickness of the film. The principal theoretical uncertainty is that the method

is a dynamical one, and the hydrodynamics of the flow of the film is not yet properly understood. However, this raises the hope that a study of the oscillations might also help to elucidate this hydrodynamics.

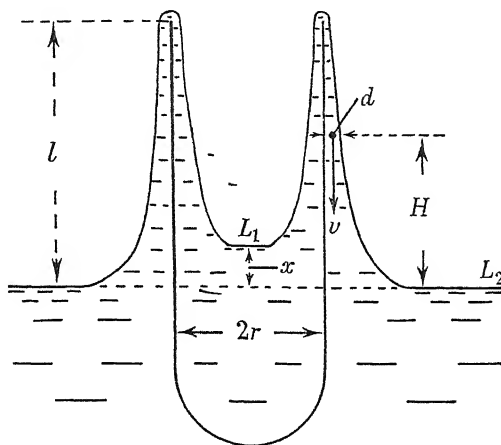


FIGURE 1. Diagram illustrating the theory of the oscillations.

(b) *Theory of the oscillations*

It is important to state explicitly the assumptions which will be made in developing the theory of the oscillations. They are:

- (1) The density and composition of the film are the same as for the bulk liquid.
- (2) The part of the film which moves is the superfluid part of density ρ_s and its motion is entirely independent of the normal part.
- (3) The shape of the film does not alter as its velocity varies.
- (4) The temperature is constant throughout the system, and so the thermo-mechanical effect may be ignored.
- (5) Dissipative effects, such as those due to viscosity, may be ignored.
- (6) The velocity in the film at a fixed height and at a given time is constant throughout the thickness of the film. If the thickness of the film varies slowly with height, this is an approximation to the assumption that the motion is irrotational ($\text{curl } \mathbf{v} = 0$).
- (7) The hydrodynamics is therefore that of an ideal classical liquid undergoing a frictionless irrotational motion.

Referring again to figure 1, if d is the thickness of the film at a height H above the free surface, v is the velocity of the film at all distances from the wall at a height H and a time t , and σ is the rate of transfer of liquid through the film at time t in $\text{cm}^3/\text{sec.}/\text{cm.}$ width of film, then the equation of continuity expresses the fact that σ must be the same at all heights, so that

$$\frac{\rho_s}{\rho} v d = \sigma = \text{a function of } t \text{ only.} \quad (1)$$

The rate of rise of the inner level in the beaker is given by

$$\rho \pi r^2 \frac{dx}{dt} = -\rho_s 2\pi r v d, \quad (2)$$

r being the radius of the beaker, x the height of the inner level above the outer level, and ρ and ρ_s the densities of the bulk liquid and superfluid component respectively. The velocity at any point in the film is consequently

$$v = -\frac{\rho}{\rho_s} \frac{r}{2} \frac{1}{d} \frac{dx}{dt}. \quad (3)$$

Euler's hydrodynamical equation

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \text{grad}) \mathbf{v} = -\frac{1}{\rho} \text{grad } p$$

takes the form

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial}{\partial z} \left(\frac{1}{2} v^2 \right) = -\frac{1}{\rho} \frac{\partial p}{\partial z}, \quad (4)$$

in which p is the pressure at a point in the liquid at a distance z from the inner level measured along the film. Substituting the value of v given by equation (3)

$$-\frac{\rho}{\rho_s} \frac{r}{2} \frac{d^2 x}{dt^2} \frac{1}{d} + \frac{1}{2} \left(\frac{\rho}{\rho_s} \frac{r}{2} \frac{dx}{dt} \right)^2 \frac{\partial}{\partial z} \left(\frac{1}{d^3} \right) = -\frac{1}{\rho} \frac{\partial p}{\partial z}. \quad (5)$$

Integrating with respect to z from the inner level L_1 to the outer level L_2 ,

$$-\frac{\rho}{\rho_s} \frac{r}{2} \frac{d^2 x}{dt^2} \left[\int_0^l \frac{dz}{d} + \int_l^{2l} \frac{dz}{d} \right] + \frac{1}{2} \left(\frac{\rho}{\rho_s} \frac{r}{2} \frac{dx}{dt} \right)^2 \left[\frac{1}{d^3} \right]_{L_1}^{L_2} = -\frac{1}{\rho} [p]_{L_1}^{L_2}. \quad (6)$$

Here l is the height of the rim of the beaker above the bath level. At L_1 and L_2 d is infinite, so the second term vanishes and the equation reduces to

$$-\frac{\rho}{\rho_s} \frac{r}{2} \frac{d^2 x}{dt^2} 2 \int_0^l \frac{dH}{d} = +gx. \quad (7)$$

The factor 2 before the integral corresponds to the fact that we have to consider the film on the outside of the beaker as well as on the inside. If, as in the actual apparatus, the radius of the outer surface is not r but R , then it is easy to show that this factor becomes $1 + r/R$. After making this alteration equation (7) represents a simple harmonic motion of period

$$\tau = 2\pi \sqrt{\left\{ \frac{\rho}{\rho_s} \frac{r}{2g} \left(1 + \frac{r}{R} \right) \int_0^l \frac{dH}{d} \right\}}. \quad (8)$$

If the variation of film thickness with height can be expressed in the form $d = k/H^n$, it follows immediately that the period will vary as l to the power $\frac{1}{2}(n+1)$ and the value of k will be

$$k = 2\pi^2 \frac{\rho}{\rho_s g} \frac{r}{2} \left(1 + \frac{r}{R} \right) \frac{l^{n+1}}{\tau^2(n+1)}. \quad (9)$$

The experiment therefore consists of a determination of the variation of the period with the height of the rim above the surface of the bulk liquid.

The amplitude of oscillation to be expected is

$$\begin{aligned} x_{\max.} &= \left(\frac{dx}{dt} \right)_{\max.} \sqrt{\left\{ \frac{\rho}{\rho_s} \frac{r}{2g} \left(1 + \frac{r}{R} \right) \int_0^l \frac{dH}{d} \right\}} \\ &= \sigma_{\max.} \sqrt{\left\{ \frac{\rho}{\rho_s} \frac{2}{rg} \left(1 + \frac{r}{R} \right) \int_0^l \frac{dH}{d} \right\}}, \end{aligned} \quad (10)$$

$\sigma_{\max.}$ being the rate of flow just before the start of the oscillations.

(c) *The velocity contour*

Certain aspects of the above theory will now be considered in more detail. First, the assumption that all layers of the film move with the same velocity is obviously of fundamental importance, since the method can, by its very nature, measure only the effective thickness of that part of the film which moves. If, as is very likely, the first two or three layers near the wall are solid, then these will make no contribution to the period. However, a continuous variation of velocity across the film is not so important, for if the velocity has the form $v = X(l) Y(d) Z(y/d)$, y being the distance from the wall, then a simple extension of the theory gives for the period

$$\tau = 2\pi \sqrt{\left(\frac{I_2}{I_1^2} \frac{\rho}{\rho_s} \frac{r}{2g} \left(1 + \frac{r}{R} \right) \int_0^1 \frac{dH}{d} \right)}, \quad (11)$$

where

$$I_1 = \int_0^1 Z(\mu) d\mu \quad \text{and} \quad I_2 = \int_0^1 [Z(\mu)]^2 d\mu.$$

It follows that the variation of period with height is not affected, but that the apparent thickness of the film is decreased by the factor I_1^2/I_2 . A careful comparison of the results of the oscillation method and a more direct method such as that of Burge & Jackson (1949) should therefore yield some information about the velocity contour in the moving film.

(d) *Thermal effects*

So far we have assumed that the motion of the film is determined only by the hydrostatic pressure and have ignored the well-known thermal effects which result from the entropy defect of the moving component. We shall now set an upper limit to these effects for the worst possible case in which the inside of the beaker is thermally isolated from the outside. As superfluid liquid flows through the film out of the beaker the inside of the beaker is heated (mechano-caloric effect) and the resulting temperature difference is equivalent to a pressure tending to force the liquid back again (thermo-mechanical effect). If ΔT is the temperature excess inside the beaker at any instant, V is the total volume of liquid inside the beaker, C is the specific heat per g. of liquid, and Q^* is the heat transferred with 1 g. of film, then

$$V\rho C \frac{d}{dt} \Delta T = -\rho\pi r^2 \frac{dx}{dt} Q^*.$$

As ΔT and x both vary sinusoidally this can be reduced to

$$VC\Delta T = -\pi r^2 x Q^*.$$

The consequent thermo-mechanical pressure head is

$$\begin{aligned} \Delta p &= -\frac{\rho Q^*}{T} \Delta T \\ &= -\frac{\rho\pi r^2 Q^{*2}}{TCV} x. \end{aligned}$$

The total pressure head, hydrostatic and thermo-mechanical, is therefore

$$\Delta p' = \rho \left(g + \frac{\pi r^2 Q^{*2}}{TCV} \right) x, \quad (12)$$

so that this thermal effect merely increases the effective value of g and so cannot influence the prediction of the film shape, though it might influence the value deduced for k . In the actual experiment r was small and V was large and the correction could not therefore have amounted to more than about 10 %, but because of the precautions taken to ensure good thermal contact between the inside and the outside of the beaker it was probably completely negligible. We shall see later that there was experimental evidence that this was so.

3. EXPERIMENTAL DETAILS

(a) *The cryostat*

The main consideration influencing the design of the cryostat was the necessity to avoid contamination of the surface holding the film by a deposit of a solidified gas, which, as Bowers & Mendelssohn (1949) have shown, can have a drastic effect on the properties of the film.† It was also desirable to avoid temperature gradients in the

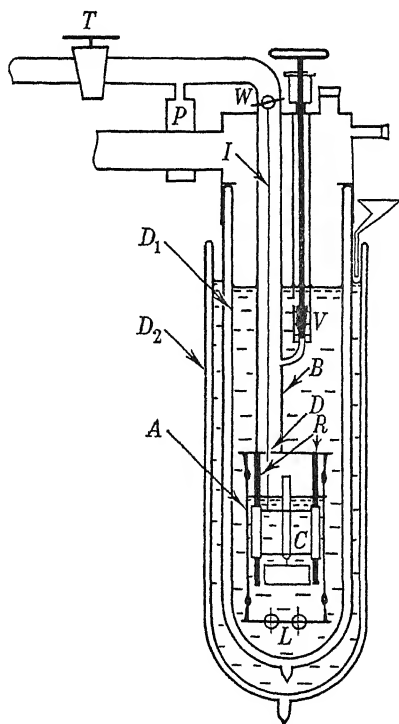


FIGURE 2. The cryostat.

neighbourhood of the film and to protect the film from radiant heat. Referring to figure 2, the experiment was performed inside the enclosure A which was completely immersed in the liquid helium contained in the flask D_1 , which, in its turn, was shielded by the liquid oxygen in D_2 . The only outlet from A was the small hole D ,

† Some early results obtained by the present author (Atkins 1948) were influenced by such deposits and gave a film thickness of the order of 10^{-5} cm. varying inversely as the height of the film.

2 mm. in diameter, which admitted only a small amount of the room radiation coming down the tube *B* and also reduced the flow of the film out of *A*, thus minimizing any thermal disturbance resulting from evaporation of this film in *B* and downward convection of the hot gas back into *A*. *A* was a 4 cm. diameter glass tube with a copper to glass seal at each end. The lower seal enabled leads to enter the enclosure when necessary via the 'Nico-seal' terminals *L*, and the upper seal enabled a soldered connexion to be made to the thin-walled german silver tube *B*. The vessel used in the experiment was mounted on the carriage *C* and was moved up and down by means of the winch *W* and the fishing line *I*, being guided by the pillars *R*. *A* was surrounded by a polished shield of tin foil with two narrow vertical slots for viewing the apparatus.

Before filling the cryostat with liquid helium, *A* was evacuated with a diffusion pump and maintained at a pressure of about 10^{-5} mm. Hg for several days. The tap *T* was then closed, and the reading of the Pirani gauge *P* noted for a sufficient time to confirm that there was no leak in the system. The cryostat was then precooled to liquid oxygen temperatures and *D*₁ was filled with liquid helium to a level well above the needle valve *V*. *V* was then opened to fill *A* with liquid helium to a convenient level and was immediately afterwards closed while still submerged in the bath liquid. It will be seen that this procedure reduced to a minimum the possibility of unwanted gas entering *A* and solidifying on the surface intended to hold the film. The arrangement had the additional advantage that, since the experiment was performed inside a closed chamber containing a fixed amount of helium, the outer level remained fixed during the flow measurements and did not, as is usually the case, fall slowly due to loss by evaporation.

(b) The vessel

Equation (10) suggests that oscillations of appreciable amplitude can be expected only when *r* is of the order of 0.3 mm., and so the beaker had to be a capillary of fairly small bore. It was in fact a soda-glass capillary of internal diameter 0.58 mm. and is labelled *X* in figure 3. It communicated via the platinum tube *Y* with the closed cylindrical metal reservoir *Z* whose floor was of thin copper foil. The purpose of *Z* was to reduce thermal effects by increasing the total volume of liquid inside the vessel (see equation (12)) and by improving the thermal contact between the inside and the outside. The main contribution to the thermal resistance of the copper foil arose from the boundary effects discovered by Kapitza (1941) and, using his numerical values, a simple calculation shows that the temperature difference due to the flow could not have risen above 10^{-6} ° K, which is negligible. The wide down-going limb *M* opening out into the bath through the constriction *N* was intended as an additional precaution against solid deposits. Any foreign gas inadvertently entering *A* and settling down as a cloud of solid droplets could not have entered *X* without first penetrating *N* and would then have found the large surface of *M* on which to settle. *M* also protected the film from convection currents of gas in *A*. The capillary *X* was calibrated along the whole of its length by means of a travelling microscope, mounted perpendicularly to the axis of *X* against the light so that the bore of *X* had the

appearance of a broad dark band. It was found to be uniform to within the accuracy of the method, which was about $\pm \frac{1}{2} \%$. The horn-shaped transition region between X and M was also calibrated in this way and an appropriate correction applied for it.

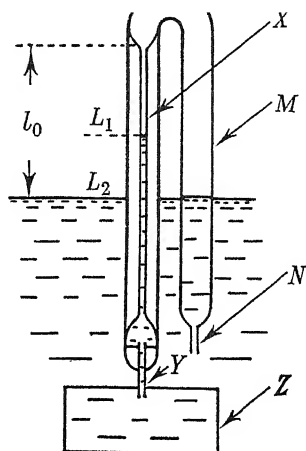


FIGURE 3. The vessel.

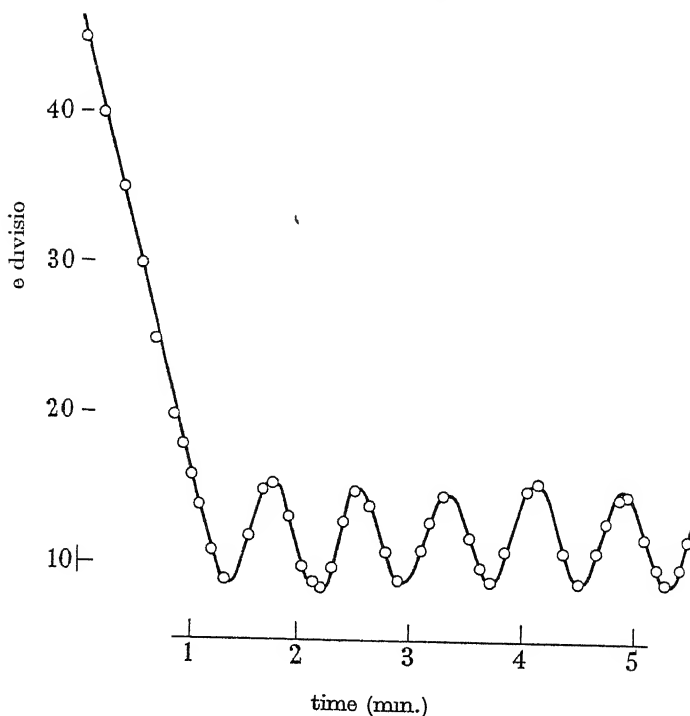


FIGURE 4. A typical oscillation (amplitude = 0.35 mm.; period = 47.2 sec.).

(c) *Experimental procedure*

The apparatus was illuminated by sodium light from which heat radiation had been filtered out by a 2 mm. thick slab of Chance ON19 glass. The meniscus was

observed through a cathetometer with a scale in its eyepiece, the sensitivity of the arrangement being a quarter of a scale division or about 0.005 cm. The inner meniscus L_1 was allowed to come to rest, and its height above the outer level L_2 was compared with the expected surface-tension rise. The absence of any detectable thermomechanical effect demonstrated that the temperature difference between the inside and the outside could not have been greater than 10^{-5}°K . The vessel was then raised or lowered and the scale reading of the inner level plotted against time as it approached its new equilibrium position and oscillated about it. A typical plot is shown in figure 4. Observation of the oscillations was made easier by the surface-tension rise, which ensured that the inner level was never obscured by the outer meniscus. Immediately after each oscillation cathetometer readings were taken for L_1 and L_2 and a fiducial mark on the outside of the capillary, so that the film height could be deduced.

4. THE EXPERIMENTAL RESULTS

(a) Variation with height

In figure 5 the logarithm of the period is plotted against the logarithm of the film height, and the plot is compared with what would be expected from the two theories and from a film of constant thickness. It is clear that the thickness decreases slightly as the height increases, but not as rapidly as either of the theories predicts. Since the experimental points lie approximately on a straight line it is possible to fit to them the curve $\tau = \tau_0 l^{\frac{1}{2}(n+1)}$ based on a film thickness of the form $d = k/H^n$. To obtain exact values of k and n two corrections must be applied to equation (8). First, it is necessary to take account of the horn-shaped region where the capillary X opens out into the wider tube M and also of the semicircular bend at the top of M . Since this correction is small it can be adequately allowed for by a small constant addition to the film height. (It should be borne in mind that the important quantity is the kinetic energy of the film and, since the film moves very slowly in the wide tube, this part of the film is far less important than the film formed on the walls of the capillary X , so that the film height is approximately the height of the top of the capillary above the bath level.) Secondly, the film thickness is determined by the height above the bath level, whereas the effective length of the film is less than this by the surface tension rise. Therefore, the experiment gives directly the variation with height above the inner meniscus and this must be converted to variation with height above the outer level. After taking these considerations into account, the exponent n becomes 0.14 ± 0.02 , and the thickness at a height of 1 cm. becomes $1.67 \pm 0.05 \times 10^{-6}$ cm., that is

$$d = \frac{1.67 \times 10^{-6}}{H^{0.14}}. \quad (13)$$

This refers to a temperature of 1.47°K and a range of heights from 0.5 to 5 cm., and, of course, is based on the assumption that the velocity is constant over the thickness of the film.

Since the rate of flow measurements, which will be discussed in part II, were made at the same time as the oscillation measurements, the result given above is adequate

for our main purpose of studying the variation of the rate of flow with the film thickness. But we must also inquire whether the variation of thickness with height given in equation (13) has any fundamental significance in connexion with the forces which form the film, or whether it is merely a consequence of the particular experimental conditions. In view of the precautions taken, it is most unlikely that the results were influenced by deposits of solidified gases, and in fact the periods were

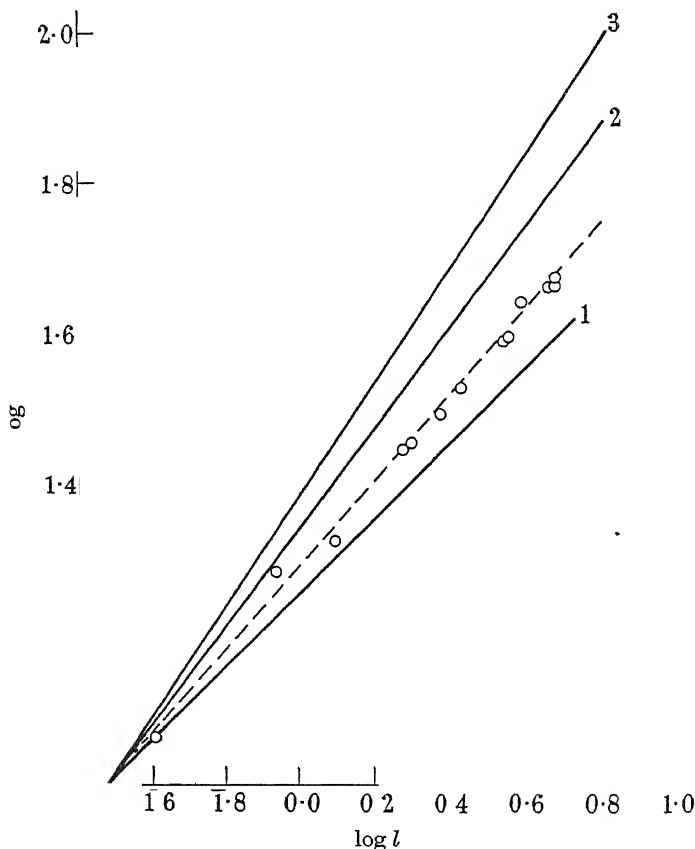


FIGURE 5. The period as a function of the height of the film.

curve	d
○	experimental
1	k_0
2	$k_s/H^{\frac{1}{2}}$ (Schiff-Frenkel theory)
3	$k_2/H^{\frac{1}{2}}$ (Bijl, de Boer, Michels theory)
Temperature = 1.47° K	

accurately reproduced to within the experimental error in four separate experiments. There is also some evidence that the film was not being influenced by incident radiation, for when a hot soldering iron was placed near the cryostat there was no measurable change in period, but there was a thermo-mechanical rise of the inner meniscus of 0.1 cm., which was at least five times as great as that due to the unavoidable 'natural' heat influx. Inconsistencies arose, however, when a vessel of

slightly different design was used. The new capillary was lead glass instead of soda glass and its internal diameter was rather smaller. It will be seen from figure 6 that the slope of the graphs $\log \tau$ against $\log l$ for the two vessels was the same within the experimental error, so that the film had the same shape in the two cases. For the lead glass capillary, however, the thickness of the film at a height of 1 cm. was 2.2×10^{-6} cm., which is significantly larger than the value of 1.67×10^{-6} cm. obtained for the soda glass capillary. This difference may be a genuine influence of the chemical nature of the surface, or it may be that the results were influenced by lack of smoothness of the glass surfaces. In the latter connexion it should be mentioned that the soda glass capillary had been cleaned with hot distilled water only, whereas the lead glass capillary had first been treated with hot chromic acid. It is clear that an exact comparison between theory and experiment cannot be made until the importance of such factors as these is more fully understood.

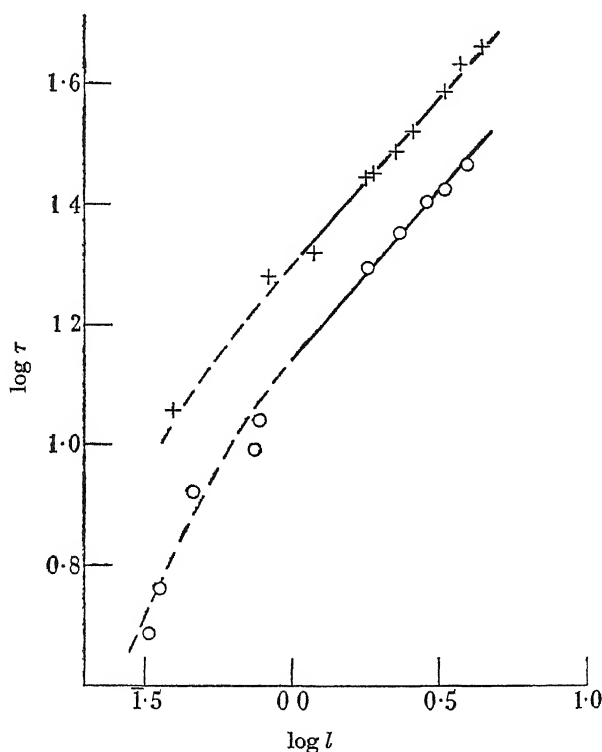


FIGURE 6. Comparison of two different vessels. +, vessel no. 1: internal diameter = 0.58 mm. O, vessel no. 2: internal diameter = 0.38 mm.

With the second vessel a more thorough investigation of the smaller film heights was made and figure 6 reveals a significant departure from linearity in this region. It is therefore very likely that the thickness cannot be represented over the whole range of heights by the formula $d = k/H^n$ and that the apparent value of n is greater at smaller values of H . The value of 0.14 for n deduced above may therefore be a coincidence depending on the range of heights used. In table 1 the present results are compared with those obtained by the optical method of Burge & Jackson (1949).

The value of n given is a mean for the stated range of heights. Since the range of heights used was different in the two cases, the two sets of observations are not necessarily inconsistent. The optical method presumably measures the total thickness of the film, whereas the oscillation method measures only the effective thickness of the moving part, and so the order of magnitude agreement between the two methods proves that a major part of the film is mobile, (see §2(c)). This disagrees with a suggestion due to Temperley (1949) that only a few of the outer layers take part in the motion.

TABLE 1

	temperature (° K)	range of H (cm.)	h (cm.) $10^{-6} \times$	n
vessel no. 1	1.47	0.5 to 5.0	1.67	0.14
vessel no. 2	1.47	0.3 to 1.0	2.2	~ 0.5
	1.47	1.0 to 4.5	2.2	~ 0.1
Burge & Jackson	1.1	0.15 to 1.5	1.9	0.28
	2.1	0.15 to 1.5		0.40

(b) *Variation with temperature*

In figure 7 the period of the oscillations has been plotted against temperature for a constant film height of 3.80 cm. To investigate the importance of the type of thermal effect discussed in §2(d), two experiments were performed with the same capillary (vessel no. 1) but different metal chambers of volume 8.35 and 0.74 cm.³ respectively. Equation (12), which assumes thermal isolation of the inside from the outside, predicts that the periods in the two cases should differ by a factor of 2.5 at

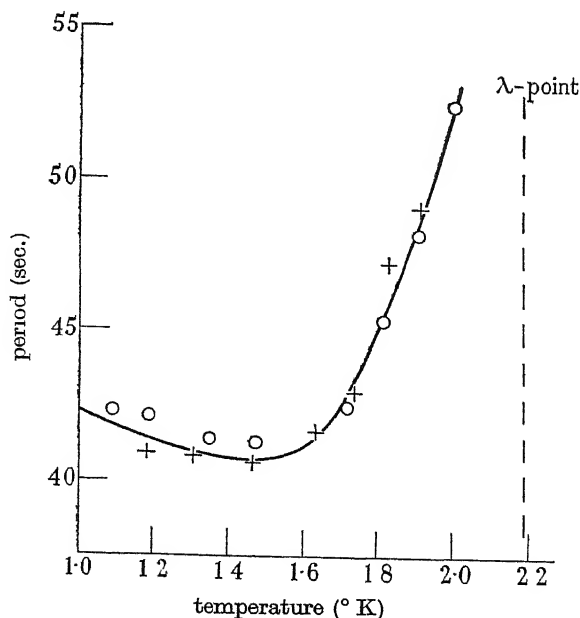


FIGURE 7. Variation of period with temperature. O, $V = 8.35$ cm.³;
+, $V = 0.74$ cm.³.

the highest temperature. The excellent agreement between the two sets of points must mean that the thermal contact between the inside and outside was very good and that it is justifiable to ignore this type of thermal effect.

The rapid increase in period above 1.6°K is probably caused by the factor ρ/ρ_s in equation (8), and is strongly in favour of the assumption that the composition of the film is similar to that of the bulk liquid, but that the liquid taking part in the superfluid motion is only a fraction of the whole. Substituting for ρ/ρ_s the values deduced by Peshkov (1946) from his second sound measurements, the curve obtained for the variation of film thickness with temperature has the form shown in figure 8. The extent of this variation is rather greater than that found by Burge & Jackson (1949). There are, however, two points in connexion with the oscillation results that need further investigation. First, it will be necessary to carry out a rather lengthy investigation to see if the functional form of the dependence of thickness on height varies with temperature. The results of Burge & Jackson (1949) suggest that there is such a variation but, unfortunately, in a direction which intensifies the discrepancy between their results and figure 8. Secondly, it is important to confirm that Peshkov's values of ρ/ρ_s are relevant to flow through very narrow channels. This can readily be done by studying the oscillations in narrow subsurface channels whose width can be assumed independent of temperature, and experiments of this type are in hand.

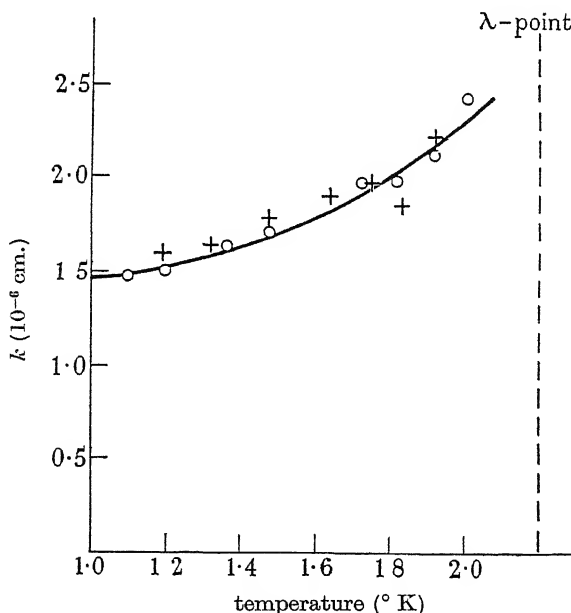


FIGURE 8. Variation of film thickness with temperature for vessel no. 1.

I should like to express my thanks to Professor J. F. Allen for suggesting this problem to me and for many helpful discussions. My thanks are also due to Mr D. V. Osborne who assisted with the measurements. During the early stages of the research I received financial assistance from the Ministry of Education and was enabled to complete the work by the award of an I.C.I. Fellowship and a Fellowship at Trinity College, Cambridge.

REFERENCES

- Allen, J. F. & Misener, A. D. 1939 *Proc. Roy. Soc. A*, **172**, 467.
 Atkins, K. R. 1948 Thesis, Cambridge.
 Bijl, A., de Boer, J. & Michels, A. 1941 *Physica*, **8**, 655.
 Bowers, R. & Mendelssohn, K. 1949 *Nature*, **163**, 870.
 Burge, E. J. & Jackson, L. C. 1949 *Nature*, **164**, 660.
 Daunt, J. G. & Mendelssohn, K. 1939 *Proc. Roy. Soc. A*, **170**, 423.
 Frenkel, J. 1940 *J. Phys. U.S.S.R.* **2**, 365.
 Kapitza, P. L. 1941 *J. Phys. U.S.S.R.* **4**, 181.
 Kikoin, A. K. & Lasarew, B. G. 1939 *Nature*, **142**, 289.
 Kistemaker, J. 1947 *Physica*, **13**, 81.
 Landau, L. 1941 *J. Phys. U.S.S.R.* **5**, 71.
 Long, E. A. & Meyer, L. 1949 *Phys. Rev.* **76**, 440.
 Peshkov, V. 1946 *Phys. Soc. Camb. Conf.* **2**, 19.
 Schiff, L. I. 1941 *Phys. Rev.* **59**, 838.
 Temperley, H. N. V. 1949 *Proc. Roy. Soc. A*, **198**, 438.
 Tisza, L. 1938 *Nature*, **141**, 913.
 Tisza, L. 1947 *Phys. Rev.* **72**, 838.

The system silver-magnesium-antimony, with reference to the theory of alloy formation

BY B. R. T. FROST AND G. V. RAYNOR

Department of Metallurgy, University of Birmingham

(Communicated by W. Hume-Rothery, F.R.S —Received 31 March 1950)

For comparison with the recently determined equilibrium diagrams of the silver-magnesium-zinc and silver-magnesium-tin alloys, the system silver-magnesium-antimony has been studied by metallographic and X-ray methods. The results are presented in the form of isothermal diagrams at 550 and 450° C. At both temperatures the intermetallic compounds Mg_3Sb_2 and AgMgSb enter into equilibrium with the primary silver-rich solid solution. At 550° C the solid alloys are confined to a composition region lying approximately between a line joining Mg_3Sb_2 to the antimony-rich side of the $(\zeta + \gamma)$ region of the silver-antimony system, and a line joining Mg_3Sb_2 to the magnesium-rich side of the β' region of the silver-magnesium system. Within this region the equilibria are similar to those at 450° C. At 450° C the compound Mg_3Sb_2 enters into equilibrium with magnesium, antimony, Mg_3Ag , the β' phase of the silver-magnesium system, and the silver-rich ternary solid solution. The compound AgMgSb enters into equilibrium with Mg_3Sb_2 , antimony, the γ and ζ phases of the silver-antimony system, and the silver-rich solid solution. The isothermal diagram, therefore, contains eight three-phase triangles, and the equilibria are dominated by the two intermetallic compounds.

The form of the equilibrium diagram is discussed in comparison with previous work, and it is shown that the effect of a compound formed by the solute elements depends on its heat of formation, which can be taken as a rough measure of its free energy. Compounds between magnesium and zinc, which have low heats of formation, have little effect on the equilibrium relations in the silver-rich alloys. Mg_3Sn , of higher heat of formation, enters into equilibrium with the β' phase of the silver-magnesium-tin system, and affects the details of the equilibrium between the $\frac{2}{3}$ electron compounds and the extent of the primary solid solubility Mg_3Sb_2 , with a still higher heat of formation, prevents equilibrium between the $\frac{2}{3}$ electron compounds, and itself enters into equilibrium with the primary ternary solid solution.

The form of the α boundary, at which both AgMgSb and Mg_3Sb_2 come into equilibrium with α , is consistent with the ideas of Hume-Rothery, according to whom the isothermal boundary at which a compound B_xC_y separates in the system ABC may be represented by the expression

$$[B]^x [C]^y = \text{constant},$$

where $[B]$ and $[C]$ are, to a first approximation, the atomic percentages of the solute elements.

1. INTRODUCTION

In previous discussion of the factors affecting the formation of ternary alloys it has been suggested (Raynor 1948*a, b*) that the effect of stable compound formation between the two solute elements is of importance in determining the nature of the equilibria which occur. Consideration of hypothetical free-energy surfaces for ternary alloys *ABC* suggests that only compounds of the solute elements (*B* and *C*) which have high heats of formation compared with those of other phases in the system may be expected to enter into equilibrium with the solvent-rich alloys; this is in general agreement with experimental data. Even when the compound does not enter into equilibrium with the primary solid solution, however, its effects on the details of the equilibrium diagram are not negligible. Thus, it has been shown (Raynor & Frost 1949) that, in the silver-magnesium-tin system, the primary solid solubility isothermals attain a maximum electron:atom ratio, for a given temperature, at a composition such that the ratio of magnesium atoms to tin atoms is 2:1. This is the ratio required to form Mg_2Sn . In the same system, the ζ phase of the silver-tin system can dissolve magnesium atoms up to a limit which again corresponds with the atomic ratio $2\text{Mg}:\text{1Sn}$. In the case of the silver-magnesium-zinc system, however, the compounds formed by magnesium and zinc have no influence upon the equilibrium relationships (Raynor & Smith 1949). This may be understood in terms of the relatively low heats of formation of $\text{Mg}_2\text{Zn}_{11}$, MgZn and MgZn_2 , compared with that of Mg_2Sn .

It was of interest to examine the form of the equilibrium relationships in the system silver-magnesium-antimony, where the heat of formation of the compound formed by the solute elements is greater than that of Mg_2Sn , the present paper reports the results of this investigation. An important consideration in the choice of this system was that relative atomic size relationships do not differ significantly from those characteristic of the previously examined silver-magnesium-tin system. The only major difference is the inclusion in the system of the compound Mg_3Sb_2 in place of the compound Mg_2Sn . Any major differences in the equilibrium relationships for the two ternary systems may, therefore, be directly attributed to the electrochemical differences between the two binary systems magnesium-tin and magnesium-antimony.

2. THE BINARY SYSTEMS

The system silver-magnesium

The equilibrium diagram for the silver-magnesium alloys is shown in figure 1, and is satisfactorily established. The silver-rich portion, up to 40 at. % of magnesium, is taken from the work of Andrews & Hume-Rothery (1943), while the remainder is based upon the diagram given by Hansen (1936) and incorporates the work of several authors. The solid solubility of magnesium in silver rises from 26.5 at. % at 300°C to 29.3 at. % at the eutectic temperature of 759.3°C. The β' phase, with a crystal structure of the caesium chloride type, has a wide range of homogeneity which narrows slightly as the temperature decreases. The phase denoted Mg_3Ag is an intermetallic compound of relatively narrow range of homogeneity; the boundaries shown in figure 1 are probable, but not fixed with certainty. The recent results of

Letner & Sidhu (1947) for the positions of the boundaries of the α and β' phases are inconsistent with figure 1, which has, however, been confirmed in the appropriate range of compositions by the present authors.

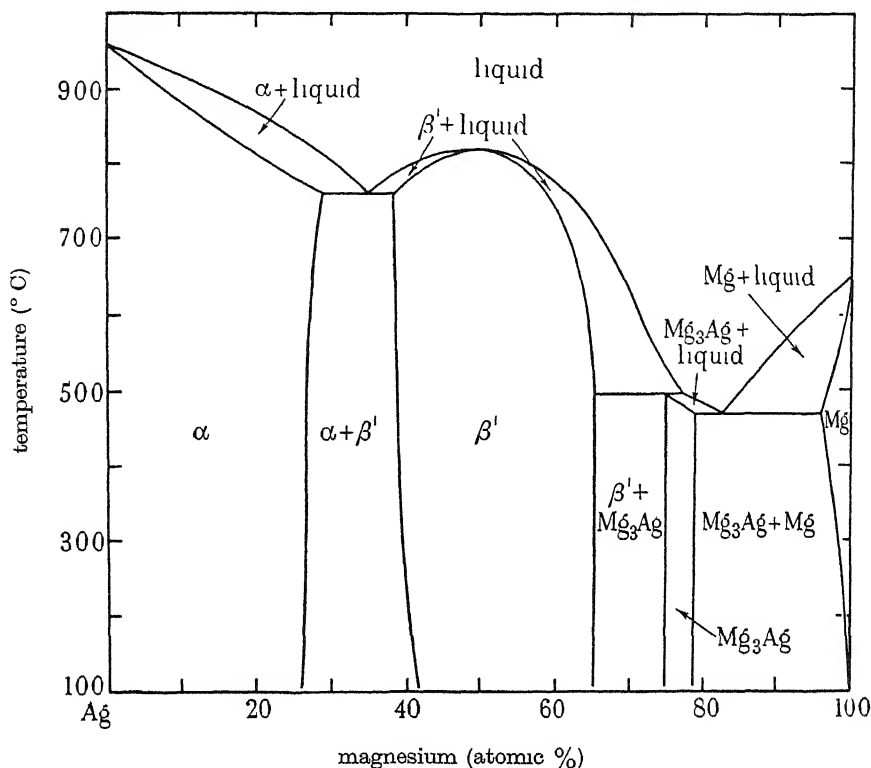


FIGURE 1. The equilibrium diagram for the silver-magnesium system.

The system silver-antimony

The silver-antimony equilibrium diagram is given in figure 2. This is also based upon that given by Hansen (1936); the silver-rich region, from 0 to 25 at. % of antimony, is drawn according to the accurate results of Hume-Rothery & Reynolds (1937). The maximum solid solubility of antimony in silver is 7.2 at. % at 702.5°C. At this temperature, α and liquid react to form the ζ phase, which is a $\frac{3}{2}$ electron compound with a close-packed hexagonal structure, extending from 8.8 to 16.3 at. % of antimony at 558.5°C, where ζ and liquid react to form the γ phase. The γ phase is a $\frac{7}{4}$ electron compound with a rhombic crystal structure.

According to Weibke & Effinger (1940), whose results have been incorporated in figure 2, the γ phase undergoes an order-disorder transformation at 440 to 449°C. Since this temperature range is below that at which the present experiments were carried out, the transformation is of little importance for the purposes of this paper.

The γ phase and antimony, in which silver is almost insoluble, crystallize from the liquid as a eutectic at 41 at. % of antimony and 486°C.

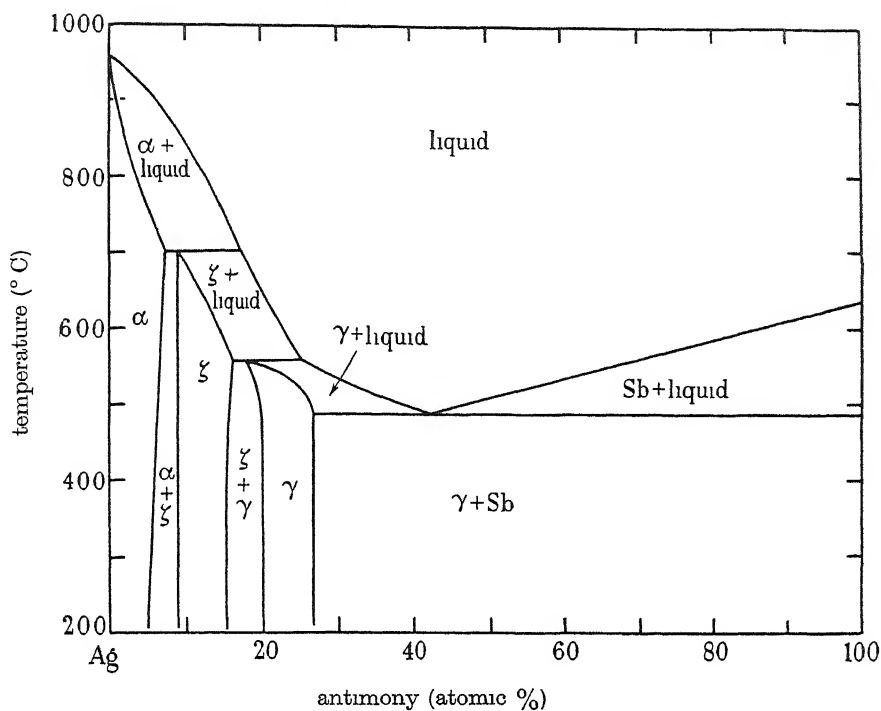


FIGURE 2. The equilibrium diagram for the silver-antimony system.

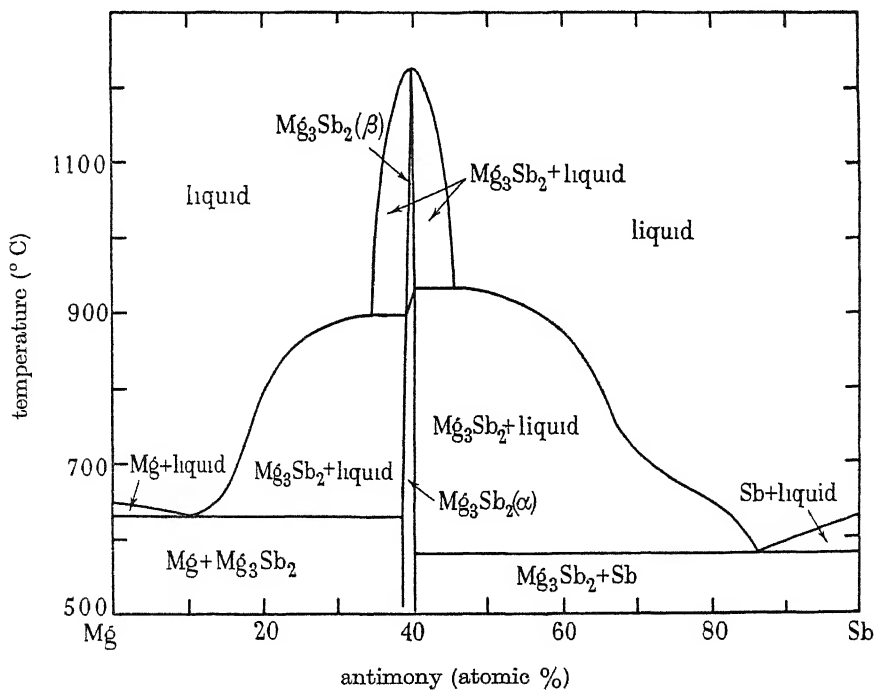


FIGURE 3. The equilibrium diagram for the magnesium-antimony system.

The system magnesium-antimony

Figure 3 shows the equilibrium diagram of the magnesium-antimony alloys (Grube & Bornhak 1934). The form of the diagram is of little importance to the present work, except for the existence and nature of the intermetallic compound Mg_3Sb_2 . This compound is stable up to 1228°C ; owing to the high electro-chemical factor in this system the reaction between magnesium and antimony is vigorous, and the heat of formation of Mg_3Sb_2 is high ($15.3\text{ kcal./g. atom}$). The compound undergoes a polymorphic transformation between 894° and 930°C , the low temperature α form has a trigonal crystal structure and is anti-isomorphous with lanthanum oxide La_2O_3 . Mg_3Sb_2 has, therefore, many of the characteristics of a salt-like compound, and has, according to Grube & Bornhak, a very narrow homogeneity range.

3. THE TERNARY SYSTEM

Before the present experiments, the only reference to the ternary system in the literature was to work by Nowotny & Sibert (1941), who proved the existence of a ternary compound AgMgSb having a crystal structure anti-isomorphous with fluorspar, CaF_2 . The heat of formation of this compound is high, but is not accurately known; it will be noted that the number of electrons per atom is $\frac{8}{3}$, as for other compounds which take the fluorspar structure.

The primary object of the research was to determine the nature of the equilibria in the silver-rich corner of the ternary model. Early experiments showed the presence, at high silver contents as well as high solute percentages, of an intermediate compound phase with properties quite different from those of the electron compounds, and the scope of the investigation was extended to include all the phases present.

The experimental methods were closely similar to those previously described in connexion with the silver-magnesium-tin system (Raynor & Frost 1949). The materials used were:

- (i) Assay silver, 99.99 % pure, supplied by Messrs Johnson, Matthey and Co. Ltd.
- (ii) Magnesium, 99.95 % pure, supplied by Messrs F. A. Hughes and Co. Ltd.
- (iii) Antimony, 99.9 % pure, supplied by the Cookson Lead and Antimony Co. Ltd.

Alloys were prepared, in 10 g. quantities, in graphite crucibles, and were well stirred under a suitable flux before casting into cold copper moulds. Annealing treatments were carried out *in vacuo*, using laboratory tube furnaces controlled at the desired temperature by Foster temperature regulators. All annealing experiments were terminated by quenching the specimens into iced water; temperatures were measured by platinum/platinum-rhodium thermocouples used in conjunction with a potentiometer. Specimens were prepared for micro-examination by exposing a new surface, and polishing in the usual manner. Micrographic characteristics are referred to below.

Confirmation of the results of the metallographic work was obtained by X-ray powder photography of selected alloys, using copper K_α radiation from a Metropolitan-Vickers demountable X-ray unit.

All critical alloys were analyzed, using as the analytical samples the actual specimens which had been examined microscopically. Several alloys were analyzed for all three components, and the totals of the percentages determined exceeded 99.9 %. It was, therefore, considered justifiable to determine only two components in the majority of cases. Of the eighty-six alloys examined only fifteen were not submitted for analysis, since their compositions were not critical for the construction of the isothermal diagrams.

In the description of experimental results, it is necessary to refer to specific alloys, which are designated in terms of their compositions. Thus 'alloy 18.30/1.61' refers to a specimen containing 18.30 at. % of magnesium and 1.61 at. % of antimony.

4. EXPERIMENTAL RESULTS

The following phases were observed in the metallographic examination:

- (i) Silver-rich solid solution, α .
- (ii) The β' phase of the silver-magnesium system.
- (iii) The Mg_3Ag phase of the silver-magnesium system.
- (iv) The ζ phase of the silver-antimony system.
- (v) The γ phase of the silver-antimony system (Ag_3Sb).
- (vi) AgMgSb .
- (vii) Mg_3Sb_2 .

The most generally useful etching reagent was a solution, in 100 ml. distilled water, of 20 g. of chromium oxide (CrO_3) and 1.5 g. of anhydrous sodium sulphate. Reaction with the ζ phase, where it existed alone, was slow, and the addition of a few drops of strong nitric acid to the etching bath improved the microstructure. Specimens were etched by immersion, and under these conditions α was revealed as yellow twinned crystals, β' as brown crystals frequently showing strain markings, and ζ as a white phase, which appeared pale blue in the presence of α or β' .

For alloys richer in antimony and magnesium, a 5 % aqueous solution of potassium dichromate was used. This reagent gave a particularly clear etch of the phase Ag_3Sb , which was coloured brown and was, therefore, easily distinguishable from the white ζ phase. In the unetched condition, distinction between these phases was difficult. The dichromate reagent was also useful in distinguishing between the three phases α , β' and Mg_3Sb_2 .

In general, Mg_3Sb_2 , AgMgSb and Mg_3Ag were highly reactive towards etching reagents, and tended to dissolve out. Alloys containing these phases were, therefore, examined before etching. It was, initially, difficult to distinguish between Mg_3Sb_2 and AgMgSb , although, when both were present in the same microstructure, the latter always appeared the lighter in colour. A 2 % alcoholic solution of picric acid, however, was found to distinguish satisfactorily between the two compounds; the X-ray method was invariably used for confirmation of these results.

The results obtained in the experimental work may be conveniently discussed in terms of the isothermal diagrams established. Alloys were examined after annealing at 450° C, and after subsequent annealing treatment at 550° C. Tests were made, by re-annealing critical alloys, to ensure that sufficient time had been allowed for the establishment of equilibrium.

(i) *The 550° C isothermal*

The results obtained on annealing specimens for 1 week at 550° C after slow heating from 200 to 600° C, at which temperature they were held for 1 day, are summarized in the isothermal diagram shown in figure 4. Both Mg_3Sb_2 and AgMgSb enter into equilibrium with the primary solid solution; in addition AgMgSb comes into equilibrium with the ζ phase and Mg_3Sb_2 with the β' phase. The region containing the

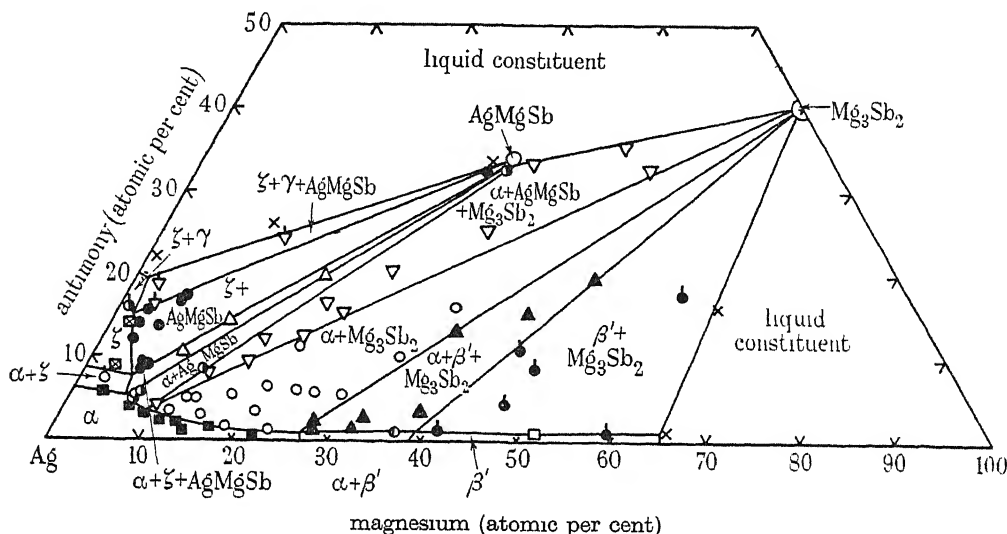


FIGURE 4. The system silver-magnesium-antimony. 550° C isothermal.

Key for figures 4, 5 and 6

α	■	$\gamma + \text{AgMgSb}$	⊖
$\alpha + \zeta$	○	$\zeta + \gamma + \text{AgMgSb}$	▽
ζ	⊠	$\zeta + \text{AgMgSb}$	●
$\zeta + \gamma$	◐	$\alpha + \zeta + \text{AgMgSb}$	△
γ	▣	$\alpha + \text{AgMgSb}$	◑
$\gamma + \text{Sb} + \text{AgMgSb}$	◈	$\alpha + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2$	▽
$\text{Sb} + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2$	◊	$\alpha + \text{Mg}_3\text{Sb}_2$	○
$\alpha + \beta'$	◐	$\alpha + \beta' + \text{Mg}_3\text{Sb}_2$	▲
β'	□	$\beta' + \text{Mg}_3\text{Sb}_2$	●
$\beta' + \text{Mg}_3\text{Ag}$	◑	$\beta' + \text{Mg}_3\text{Sb}_2 + \text{Mg}_3\text{Ag}$	▼
contained chilled liquid	×		

solid alloys is bounded by the magnesium-rich boundary of the ($\beta' + \text{Mg}_3\text{Sb}_2$) phase-field, by the antimony-rich boundary of the ($\zeta + \gamma + \text{AgMgSb}$) field, and the antimony-rich boundary of the ($\text{AgMgSb} + \text{Mg}_3\text{Sb}_2$) region. Outside these limits, the alloys contain the liquid phase.

The limit of primary solid solubility is defined to a good degree of accuracy by the alloys plotted. On the silver-magnesium axis, the limit occurs at 27 at. % at 550° C. The boundary meets the corner of the ($\alpha + \beta' + \text{Mg}_3\text{Sb}_2$) field at an antimony content of 1 at. %. This antimony content remains almost constant while the magnesium content falls to 20 at. %. Alloy 18.30/1.61 contained only a small amount of Mg_3Sb_2 , and lies just within the two-phase field. Alloys 11.35/3.45 and 9.90/4.11 also contained small quantities of Mg_3Sb_2 , the latter contained AgMgSb also. Alloys 6.95/4.91 and 6.94/5.41 contained small amounts of AgMgSb . The $\alpha/(\alpha + \text{AgMgSb})$ boundary is fixed at both ends by the corners of carefully established three-phase triangles; the single-phase alloy 7.00/4.00 limits the penetration of the boundary into the homogeneous α region. The amount of AgMgSb in alloy 6.95/4.9, however, suggests that the $\alpha/(\alpha + \text{AgMgSb})$ boundary is convex towards the silver-rich corner of the diagram. The maximum primary solid solubility occurs at the corner of the ($\alpha + \zeta + \text{AgMgSb}$) triangle, from this point, the $\alpha/(\alpha + \zeta)$ boundary is drawn as a straight line to the limit of solid solubility of antimony in silver, being approximately bracketed by the alloys 3.26/5.90 and 2.53/7.38. It will be appreciated that the ($\alpha + \zeta + \text{AgMgSb}$)/($\alpha + \text{AgMgSb}$) boundary is accurately located by alloys 20.07/19.50 and 32.35/32.73. The intersection of this boundary with the $\alpha/(\alpha + \zeta)$ boundary defines the corner of the ($\alpha + \zeta + \text{AgMgSb}$) triangle. Thus, any deviation of the $\alpha/(\alpha + \zeta)$ boundary to the antimony-rich side of the position shown in figure 4 would result in the corner of the three-phase triangle lying at a composition inconsistent with the amount of AgMgSb in alloy 6.95/4.91. The amounts of the α and ζ phases present in alloy 2.53/7.38 were consistent with the phase boundaries drawn in figure 4, which are supported by the evidence of the other boundaries. The alloy 3.00/9.00 directs the ($\alpha + \zeta$)/ ζ boundary, while the $\zeta/(\zeta + \text{AgMgSb})$ boundary is bracketed by alloys 3.25/12.05 and 2.00/14.00; the position of the latter boundary as drawn is supported by the amounts of AgMgSb in alloys 5.91/8.49, 6.41/9.12, 5.88/9.43, 3.00/14.00 and 3.14/15.73. The ternary ζ field is limited by the ($\zeta + \gamma$) alloy 1.00/16.00, and the ($\zeta + \gamma$) field, as indicated by the amounts of (AgMgSb) in alloys 3.80/16.00 and 2.98/18.78, is narrow. The remaining phase fields are adequately fixed by the alloys plotted; it will be noted that the three alloys placed in the narrow ($\alpha + \zeta + \text{AgMgSb}$) triangle taken together with the two-phase alloys in adjacent fields, determine its shape and direction accurately.

The recognition of both AgMgSb and Mg_3Sb_2 in the ($\alpha + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2$) triangle was at first difficult, but with careful polishing and etching, satisfactory identifications were possible, and were confirmed by subsequent X-ray diffraction patterns. The presence of this field affords the clearest indication of the nature of the equilibria in the system, and it is adequately located by the alloys examined. The limits of the ($\alpha + \text{Mg}_3\text{Sb}_2$) and ($\alpha + \beta' + \text{Mg}_3\text{Sb}_2$) fields are also well defined, and the evidence obtained from all the alloys examined admits of no other interpretation than that given in figure 4.

of this field implies equilibrium between Mg_3Sb_2 and Mg_3Ag , so that the equilibrium relationships in the magnesium-rich corner of the ternary diagram may readily be inferred.

It is of interest to note that the structures of alloys 15.00/50.00 and 35.00/50.00, respectively ($\gamma + \text{AgMgSb} + \text{Sb}$) and ($\text{Mg}_3\text{Sb}_2 + \text{AgMgSb} + \text{Sb}$), allow the form of the diagram in the antimony-rich region to be drawn. The micro-structures of these alloys also give indirect evidence with regard to the position in the diagram of the AgMgSb phase. According to Nowotny & Sibert (1941), this phase lies slightly to the antimony-rich side of a straight line joining the Mg_3Sb_2 and γ phases. If it were to contain less antimony, there would be a strong probability that Mg_3Sb_2 would enter into equilibrium with the γ phase, as it does with the other binary phases. The fact that this equilibrium does not occur is in favour of the position of the AgMgSb phase plotted in figure 5, further support is given by the microstructures of the four alloys with compositions in the neighbourhood of the compound.

As at 550°C , the compound Mg_3Sb_2 dominates the equilibria to a great extent, and enters into equilibrium with α , β' , Mg_3Ag , Mg and Sb . The β' phase and the ζ phase enter respectively into equilibrium with α and Mg_3Ag , and with α and γ . These equilibria involve eight three-phase triangles, as shown in figure 5.

5. X-RAY EXAMINATION

In order to assist in the correct identification of phases seen under the microscope, diffraction patterns were obtained from a number of alloys. For this purpose, annealed alloys were filed; the unsieved filings were re-annealed at the temperature from which the lump was quenched, and rapidly cooled. Standard photographs of the diffraction patterns of all the binary phases were obtained for comparison with those from ternary alloys, while the diffraction patterns from the alloys lying close to AgMgSb served to characterize this phase, for which a fluorspar structure was confirmed.

The X-ray examination of alloys lying well within the ($\alpha + \text{Mg}_3\text{Sb}_2$) field confirmed the existence of this equilibrium at an early stage in the work, and subsequent experiments entirely supported the general interpretation of the micrographical results.

6. DISCUSSION

The results of the micrographic and X-ray examination of the silver-magnesium-antimony system show that the equilibria are dominated by the stable compound Mg_3Sb_2 , which enters into equilibrium with the primary silver-rich solid solution. The ternary compound AgMgSb also comes into equilibrium with the solid solution, and there must be a discontinuity at the corner of the ($\alpha + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2$) triangle such that the boundaries $\alpha/(\alpha + \text{AgMgSb})$ and $\alpha/(\alpha + \text{Mg}_3\text{Sb}_2)$, if produced, lie in the two-phase regions and not in the homogeneous area. This fact has been utilized in drawing figures 4 and 5, and it is of interest to examine the resulting α boundary, the experimental evidence for which is presented in § 4. Figure 6 shows, on an enlarged scale, the most probable position for the α boundary at 550°C up to 25 at. % of magnesium; the lines have been drawn with due regard to the amounts of

the microconstituents present in the two-phase and three-phase alloys, and it is estimated that the boundary can hardly differ from that shown by more than ± 0.15 at. % of antimony at a given percentage of magnesium. In figure 7 the boundary has been plotted as $\log [\text{Mg}]$ against $\log [\text{Sb}]$ where $[\text{Mg}]$ and $[\text{Sb}]$ are the atomic

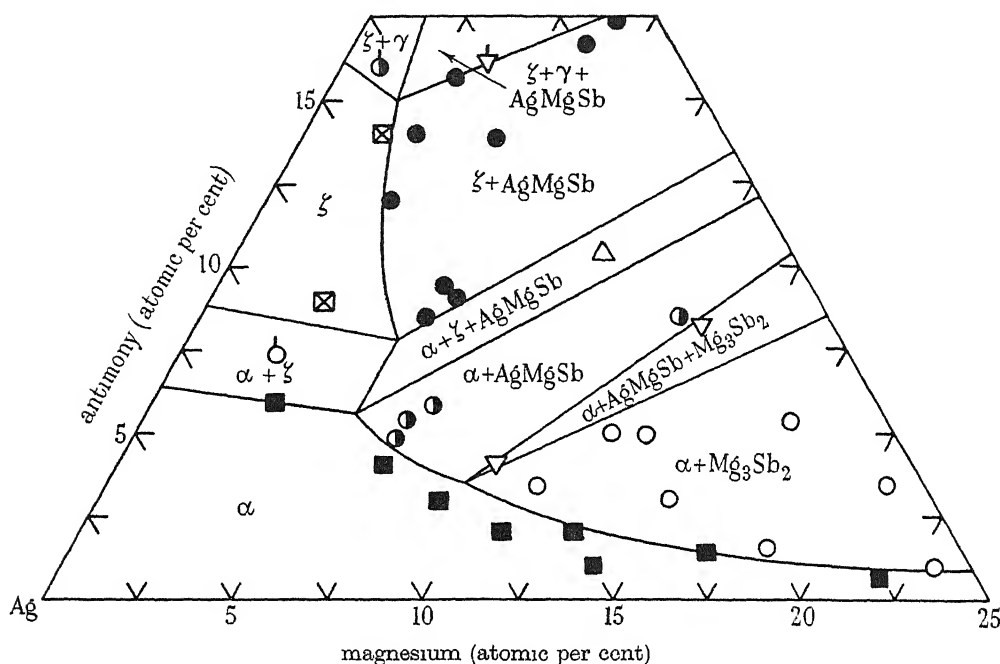


FIGURE 6. The α boundary of the system silver-magnesium-antimony at 550°C .

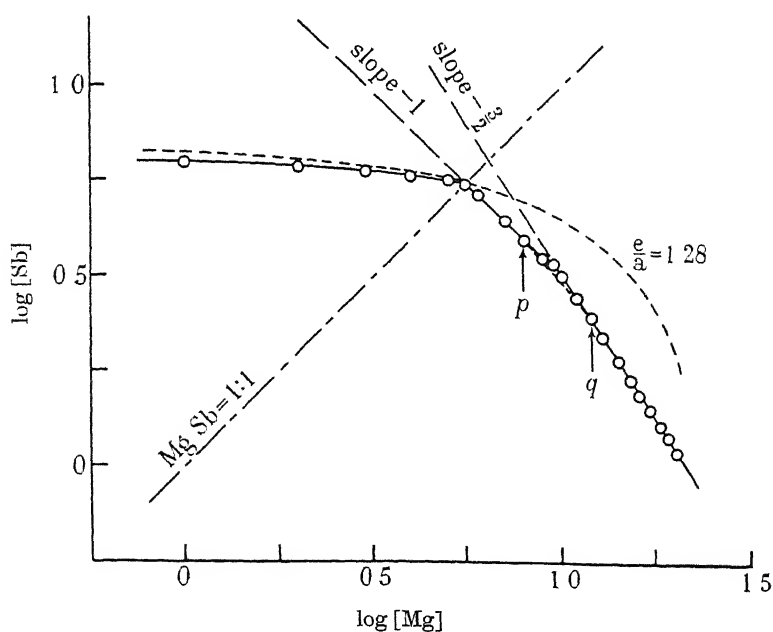


FIGURE 7. The α boundary at 550°C plotted as $\log [\text{Mg}]$ against $\log [\text{Sb}]$.

percentages of the two solutes; on the same diagram, a line of constant electron:atom ratio of 1.28 has also been drawn. Starting from the $\alpha/(\alpha + \zeta)$ boundary of the silver-antimony system, the boundary, over the portion where equilibrium between α and an electron compound is concerned, is consistent with the maintenance of an almost constant electron:atom ratio. At the corner of the $(\alpha + \zeta + \text{AgMgSb})$ triangle there is a change in direction and the experimental points for the $\alpha/(\alpha + \text{AgMgSb})$ boundary are consistent with a slope of -1 . Over the $\alpha/(\alpha + \text{Mg}_3\text{Sb}_2)$ boundary, this slope is not maintained; instead, the experimental points lie close to a straight line of slope $-\frac{3}{2}$. The relatively sharp change of slope from -1 to $-\frac{3}{2}$ is due to the discontinuity at the corner of the $(\alpha + \text{AgMg} + \text{Mg}_3\text{Sb}_2)$ triangle; if this discontinuity were absent, and the two branches of the solubility isothermal joined up smoothly, the logarithmic plot between 8% and 12% of magnesium would follow the broken curve pq . The main conclusion that the $\alpha/(\alpha + \text{AgMgSb})$ and $\alpha/(\alpha + \text{Mg}_3\text{Sb}_2)$ boundaries are consistent with straight line relationships remains unaffected for the greater part of the two boundaries. It is to be noted that where the primary solid solution is in equilibrium with Mg_3Sb_2 , there is no tendency to maintain a constant electron:atom ratio. Where equilibrium with a compound is involved, therefore, there is a clear tendency for the solubility curve to conform to the general relationship

$$[B]^x [C]^y = K, \quad (1)$$

where the compound precipitating has the formula B_xC_y . This is in conformity with the conclusions of Hume-Rothery (1936), who studied the primary solubility curve at which Mg_2Si separates in the system aluminium-magnesium-silicon, and found that the relationship

$$[\text{Mg}]^2 [\text{Si}] = K \quad (2)$$

was obeyed. Similar effects were observed in the system silver-magnesium-tin along the $\beta'/(\beta' + \text{Mg}_2\text{Sn})$ boundary, for which $\log [\text{Mg}]$ plotted against $\log [\text{Sn}]$ gave a straight line of slope $-\frac{1}{2}$.

In the silver-magnesium-tin system, it was found that the maximum electron:atom ratio of the α boundary, and the whole of the $\zeta/(\zeta + \beta')$ boundary, occurred at the atomic ratio of 2Mg:1Sn, this being the atomic ratio required for the formation of Mg_2Sn . In the present case, the primary solid solution remains in equilibrium with the ζ electron compound, and the electron:atom ratio rises slowly from 1.25 to a maximum of 1.28, until the atomic ratio of 1Mg:1Sb is reached. At this stage, it enters into equilibrium with AgMgSb , and the electron:atom ratio begins to fall, so that in this system also the presence of compounds exerts an influence on the details of the equilibria. The whole of the $(\alpha + \text{AgMgSb} + \zeta)/(\alpha + \text{AgMgSb})$ boundary closely follows the 1Mg:1Sb ratio.

Along the $\alpha/(\alpha + \text{AgMgSb})$ boundary, the ratio of magnesium to antimony atoms increases, and considerably exceeds the ratio 3Mg:2Sb before equilibrium with Mg_3Sb_2 occurs. This is, however, not unexpected, since if the logarithmic relationships are obeyed by the two boundaries, the point of the $(\alpha + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2)$ three-phase triangle occurs, to a first approximation, at such a composition that the following equations are simultaneously satisfied:

$$\log [\text{Mg}] + \log [\text{Sb}] = \log K_1, \quad (3)$$

$$3 \log [\text{Mg}] + 2 \log [\text{Sb}] = \log K_2. \quad (4)$$

The solution depends upon the relative values of K_1 and K_2 . In this case, taking as fixed points the corner of the $(\alpha + \zeta + \text{AgMgSb})$ triangle and the point on the $(\alpha + \text{Mg}_3\text{Sb}_2)$ boundary at 20/1.1, it may be calculated that the corner of the $(\alpha + \text{AgMgSb} + \text{Mg}_3\text{Sb}_2)$ triangle should lie close to the composition 9.7/3.3, which is consistent with the experimental results summarized in figure 6.

The ζ phase of the silver-tin system will dissolve magnesium until an atomic ratio of 2Mg:1Sn is reached, at which stage it enters into equilibrium with the β' phase of the silver-magnesium system. In the present case, simple equilibrium between the $\frac{3}{2}$ electron compounds is prevented by the presence of the stable compounds. The type of equilibrium diagram is changed completely from that of the silver-magnesium-tin alloys, and the β' phase dissolves less than 1 at % of antimony. The maximum solubility of magnesium in the hexagonal ζ phase is 5.5 at. %.

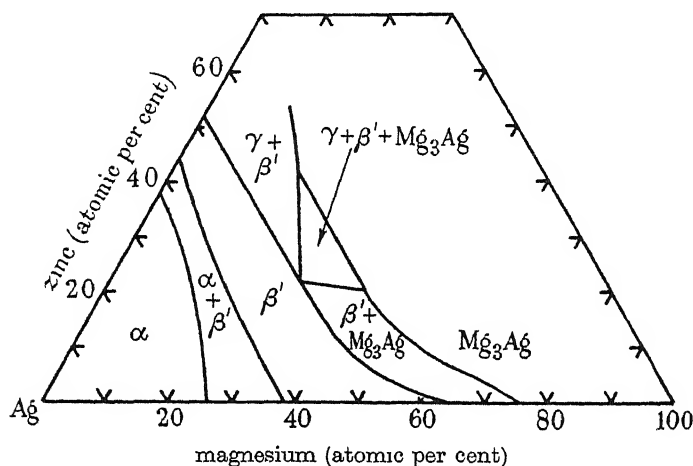


FIGURE 8. The system silver-magnesium-zinc.

It is of interest to compare the isothermal diagrams, at a temperature of 450°C, for the three systems recently studied in this laboratory. Figure 8 shows the silver-magnesium-zinc system, in which the magnesium-zinc compounds do not enter into equilibrium with the $\frac{3}{2}$ electron compounds. The maximum heat of formation in the magnesium-zinc system is 4.2 kcal./g. atom for MgZn_2 (Weibke & Kubaschewski 1943). In the magnesium-tin system, the compound Mg_2Sn has a heat of formation of 6.1 kcal/g.atom (Kubaschewski & Walter 1939), and in the corresponding ternary system, this phase enters into equilibrium with the $\frac{3}{2}$ electron compound β' (figure 9). Mg_3Sb_2 has the much higher heat of formation of 15.3 kcal./g.atom (Kubaschewski & Walter 1939), and as shown in figure 10, prevents equilibrium between the $\frac{3}{2}$ electron compounds and itself enters into equilibrium with the primary solid solution. Although, therefore, the systems silver-magnesium-tin and silver-magnesium-antimony are very similar with regard to alloying factors such as the size factor, the difference between the electrochemical factors in the systems magnesium-tin and magnesium-antimony leads to a complete change of type. This comparison forms a striking example of the part played by stable compound formation in ternary alloy systems.

It has already been shown (Raynor 1948*b*; Raynor & Frost 1949) that the influence of compound formation can be interpreted in terms of the relative heats of formation of the phases which occur in a ternary system, the heat of formation being taken as an approximate measure of the free energy. It is not possible to apply this conception quantitatively in the present case, since only heats of mixing are known for the silver-magnesium and silver-antimony systems. It is possible, however, to interpret the difference between the equilibrium diagrams of figures 9 and 10. The equilibria of figure 9 indicate that the free-energy surfaces for the silver-magnesium-tin system,

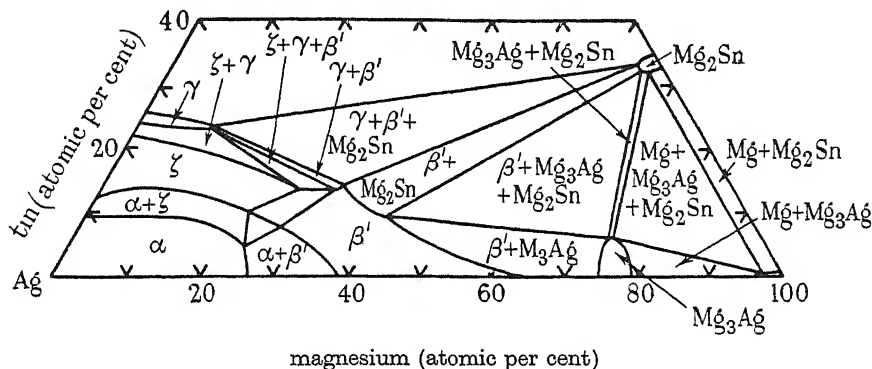


FIGURE 9. The system silver-magnesium-tin.

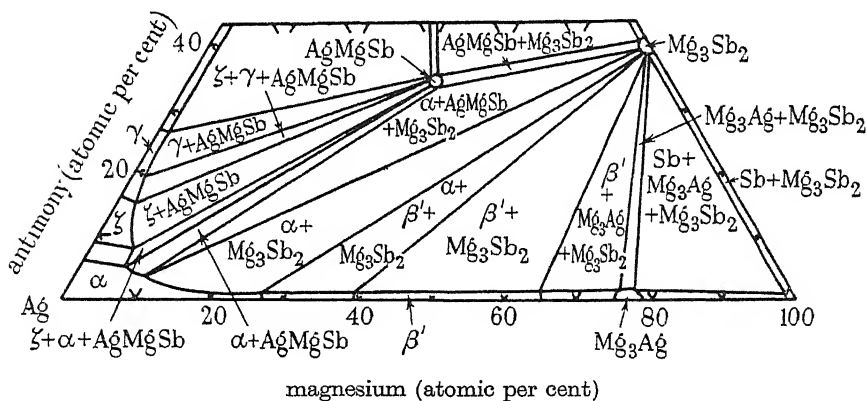


FIGURE 10. The system silver-magnesium-antimony.

and the tangent planes which determine the three-phase equilibria, must be disposed somewhat as shown diagrammatically in figure 11, where, for clarity, free energy has been drawn increasing in the downward direction. If now the free energy of the compound between the solutes is markedly decreased, so that the corresponding free energy peak in the model is enhanced, a similar form of equilibrium to that in figure 11 would give rise to the free energy model shown in figure 12. This, however, is inadmissible, as a re-entrant region is introduced, so that in certain composition regions a lower free energy can be obtained by an alternative arrangement of the tangent planes, as in figure 13, which necessitates equilibrium between the compound formed by the solutes and the solvent-rich solid solution.

It is clear, therefore, that only in the absence of stable compounds, of high heat of formation, between the solute elements can simple equilibria be established between the $\frac{3}{2}$ electron compounds of the silver alloys. If the compound involved has an insufficiently high heat of formation to enter into equilibrium with the primary solid

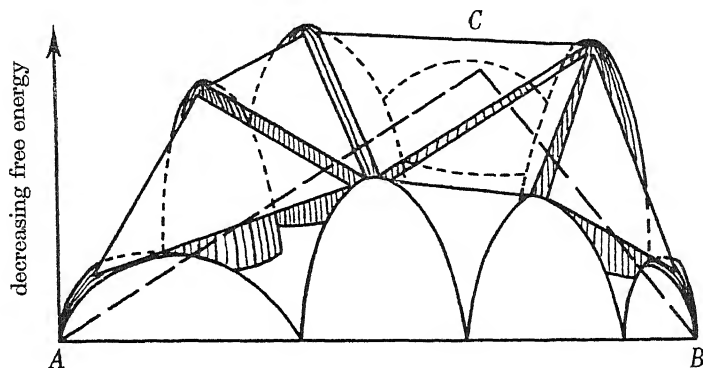


FIGURE 11. Probable form of the free-energy surfaces for the equilibria of figure 9.

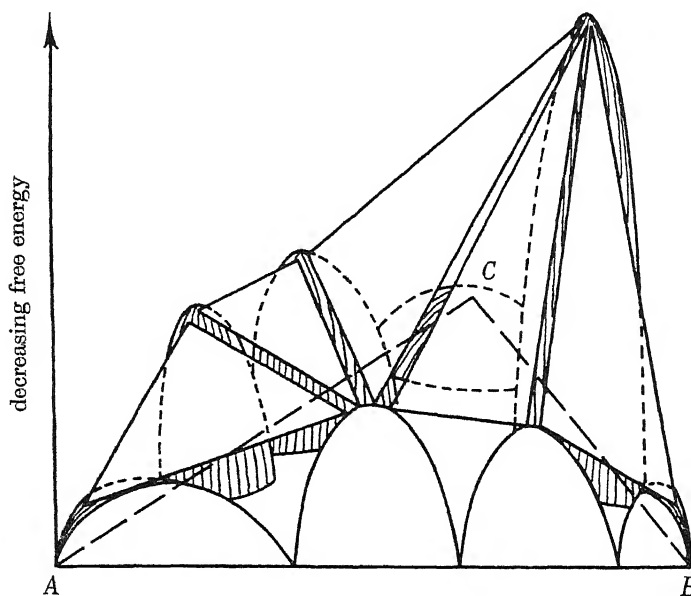


FIGURE 12. Inadmissible distribution of tangent planes in the presence of a very stable compound.

solution, it may still influence the details of the equilibrium diagram, as in the silver-magnesium-tin system. Compounds of low heat of formation have very little effect on the silver-rich alloys, and allow the size and electronic factors to operate without interference.

This research was carried out in the Metallurgy Department of the University of Birmingham, under the general direction of Professor D. Hanson, to whom the

authors' thanks are due for his interest and support. The authors must also acknowledge the valuable assistance of Mr G. Welsh in the experimental work, and the analytical work of Mr A. J. Hawkes.

The authors express their gratitude to the Department of Scientific and Industrial Research, the Royal Society, the Chemical Society, and Imperial Chemical Industries Ltd., for generous financial assistance.

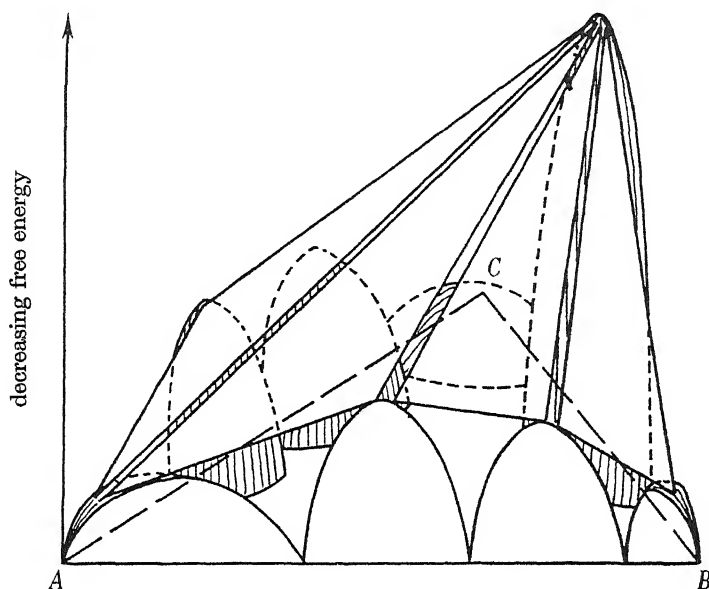


FIGURE 13. Permissible distribution of tangent planes in the presence of a very stable compound.

REFERENCES

- Andrews, K. W. & Hume-Rothery, W. 1943 *J. Inst. Met.* **69**, 485.
 Grube, G. & Bornhak, R. 1934 *Z. Elektrochem.* **40**, 140.
 Hansen, M. 1936 *Der Aufbau der Zweistofflegierungen*. Berlin: Julius Springer.
 Hume-Rothery, W. 1936 *Phil. Mag.* [vii] **22**, 1013.
 Hume-Rothery, W. & Reynolds, P. W. 1937 *J. Inst. Met.* **60**, 365.
 Kubaschewski, O. & Walter, A. 1939 *Z. Elektrochem.* **45**, 630, 732.
 Letner, H. R. & Sidhu, S. S. 1947 *J. Appl. Phys.* **18**, (9), 833.
 Nowotny, H. & Sibert, W. 1941 *Z. Metallk.* **33**, 391.
 Raynor, G. V. 1948a *Phil. Mag.* [vii], **39**, 212.
 Raynor, G. V. 1948b *Phil. Mag.* [vii], **39**, 218.
 Raynor, G. V. & Frost, B. R. T. 1949 *J. Inst. Met.* **75**, 777.
 Raynor, G. V. & Smith, R. A. 1949 *J. Inst. Met.* **76**, 389.
 Weibke, F. & Effinger, I. 1940 *Z. Elektrochem.* **46**, 53, 61.
 Weibke, F. & Kubaschewski, O. 1943 *Thermochemie der Legierungen*. Berlin: Julius Springer.

The fine structure of atmospheric turbulence in relation to the propagation of sound over the ground

By E. G. RICHARDSON, *King's College, Newcastle upon Tyne*

(Communicated by O. G. Sutton, F.R.S.—Received 3 January 1950—
Revised 19 April 1950)

[Plates 2 and 3]

With the aid of hot-wire instruments, measurements have been made of the fluctuations of the wind in the lowest 50 ft. of the atmosphere and, in particular, of the mean eddy diameter and of the ratio of vertical and cross-wind components in such fluctuations. It appears that the turbulent elements near the ground are smaller in size and more nearly isotropic in character than those higher up. Some correlation between the temperature and velocity fluctuations near the ground has been observed.

Simultaneous measurements of the fluctuations in sound intensity at a distance from a steady source and of the intervening gustiness have been made in order to discuss the relation between sound scattering and background noise, on the one hand, and the intensity and scale of the prevailing turbulence on the other. The effect of the latter on the phase relationships between the signals picked up at two points is also demonstrated.

A considerable amount of experimental research in recent years has been devoted to the fluctuations of the atmosphere, apparent in local changes of pressure, or of the velocity and direction of the wind. As a meteorological view embraces the atmosphere as a whole, such studies are of a macrocosmic nature, both in respect of space and time. There is, however, a growing tendency to think in terms of a smaller scale and to measure over a restricted region and for a short period of time. Data of this type, though not yet numerous, have been collated in terms of the eddy viscosity and transport theories of L. F. Richardson and G. I. Taylor. Sutton (1934) has had notable success in applying such ideas to atmospheric diffusion on a moderate scale. While such data, taken with simultaneous measurements of the attenuation of sound, are of use in explaining the *average* loss which open-air sound suffers in a period of minutes, the scale is still too coarse to explain the violent fluctuations in sound level which may be picked up at a moderate distance from a source of constant output under usual atmospheric conditions. The inertia of the instruments used heretofore in the atmosphere is such that much of those fluctuations in the conditions which are so important for the propagation of sound pass unnoticed. There seems to be little information about what one might call the fine structure of atmospheric turbulence in its relation to the more generally recorded mean meteorological data.

It was the main object of this work to make suitable records of these meteorological fluctuations, covering a range of conditions, and to correlate such records with simultaneous records of sound propagation in the locality. It was also hoped to be able to use sound recording as an alternative and perhaps less tedious means of deducing local eddy viscosities and diffusivities than those hitherto used.

The work began in 1940 and had soon to be abandoned. It was resumed in 1945 and has continued since.

PART I. MEASUREMENT OF WIND FLUCTUATIONS

Fluctuations in the meteorological elements may be considered under the two aspects: (1) the fluctuations at one place may be studied in relation to the conditions prevailing in the vicinity; (2) the fluctuations at two nearby places may be compared for the purpose of delineating a correlation coefficient between them or a 'mean eddy diameter'. For 'gustiness' on a rather larger space and time scale than that envisaged in this research, data have usually been derived from continuous records of a Dines anemometer. Discussions of these have been given by Scrase (1930), Best (1935) and Carruthers (1943).

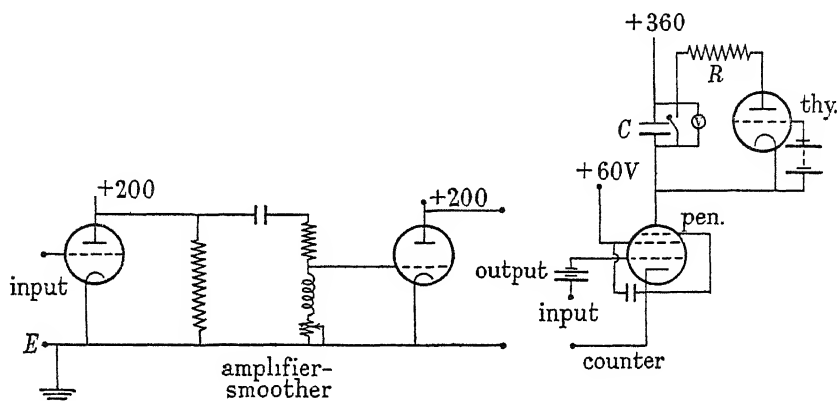


FIGURE 1. Turbulence meter, amplifier and counter circuits.

Two sites were chosen for the research now to be described. Neither of them could be described as ideal in the meteorological sense, but as time passes it seems less and less likely that the meteorologist's ideal will ever be attained, so that one must not do more than claim that a series of observations, collected over a long period of time, is typical of the site chosen. The first site, at which most records were obtained—over a period of four years—was Durham Observatory, which lies on an eminence and afforded the roof of the dome (50 ft. high) and several grassy fields round about. The second was a spot on the desert plain adjoining Cairo, from which a travelling laboratory brought records covering three summers.

At Durham, three thin constantan-copper thermocouples were mounted in horizontal tubes 3 in. diameter and 1 ft. long, to protect them from the sun and rain, at heights of 0.5, 4 and 48 ft. respectively, the low ones being occasionally fanned in hot weather before taking readings. At Cairo, Assmann psychrometers at heights of 10 and 60 cm. were set up and the difference between the aspirated dry-bulb readings taken.

The following pieces of apparatus were used for delineating the local dynamic state of the atmosphere: (1) a 'turbulence meter' by which the local intensity of turbulence was measured, amplified, compensated for frequency and rectified (figure 1). When the rectified potential on the $6\mu\text{F.}$ condenser reaches 400 V, it discharges through a thyatron and moves a counter one digit. The output of the meter is then proportional to u' , the r.m.s. value of the instantaneous deviation from the mean wind U in that direction, u, v, w being the three component fluctuations in the x, y, z directions;

(2) an apparatus for estimating the down-wind correlation of u fluctuations or 'mean eddy diameter'; (3) apparatus employing hot-wires or smoke diffusion to compare the cross-wind and vertical fluctuations v and w near the ground.

The wires used were of platinum 0.001 in. diameter by $\frac{1}{2}$ in. long, heated by 0.2 A or 0.0005 by $\frac{1}{4}$ in. long heated by 0.05 A. Owing to the scale of the phenomenon, shorter and finer wires are unnecessary and liable to fracture out-of-doors. For measurements near the ground the fork was attached to a demountable aerial; another was fixed just below the wind vane of the Observatory at a height of 53 ft. above the ground.

Using the meter, the number recorded on the counter in 1 min. could be used as a measure of the intensity of turbulence in the light of a calibration made in the laboratory, performed by oscillating a similar wire on a vibrator whose frequency and amplitude could be varied. Tests made over periods somewhat longer and shorter than 1 min. showed that conditions were fairly established in that period, but in large lapse rates and light winds, conditions might change so quickly that it was more difficult to establish a characteristic u/U . This is shown by the spread of points at the large lapse end of figure 3.

Measurement of mean eddy diameter

Some indication of the scale of turbulence was obtained by Gödicke (1935) and Findeisen (1936) using identical apparatus. This consisted of two platinum hot-wires mounted down-wind or across-wind at distances apart varying from less than 1 cm. up to 60 cm. After amplifying the potentials across the two wires these were fed into a Wheatstone bridge, so that the galvanometer in the bridge gave a measure of the mean of the instantaneous difference of velocity $U_A - U_B$ at the two stations.

In this present work it was considered preferable to compare the *fluctuation* of velocity at two points as to magnitude and phase, either by exhibiting the *fluctuations* in the velocity at A and B on a cathode-ray oscillograph, as Prandtl & Reichardt (1933) did for flow in wind tunnels, or by calculating the coefficient of correlation R , which is equal to $\overline{u_A u_B} / \mu'_A \mu'_B$, from an apparatus such as that described below.

It is also possible to gain information with regard to mean eddy size by analyzing the wave form of the fluctuations of a single hot-wire in respect of relative amplitudes of frequency components; these can then be assigned to the corresponding eddy lengths since the mean speed is known. This operation is not easy to perform in every case. One could distinguish three types of records:

(a) almost periodic or harmonic variations persisting unchanged for periods of less than 1 min., probably indicating the presence of Bénard or cellular vortices (figure 9, plate 2);

(b) bursts of unsteadiness rapidly relaxing to rise again, the time of relaxation being quasi-periodic (figure 10, plate 2). Both (a) and (b) occurred in 'thermals';

(c) very gusty incoherence in which all sizes of eddies, decreasing in amplitude down to a low limit of size, were prominent (figure 13, plate 2). These were characteristic of windy weather at zero or normal lapse rate.

Figure 2 shows the variation of u' with U over all lapse rates. In spite of considerable scatter of points, there seems to be a linear relationship among the Durham results

which mostly refer to small lapse rates. When gustiness is plotted against lapse rate figure 3 results.

Prandtl & Reichardt (1933) and others since have used an electrodymanometer to measure R in the flow in wind tunnels. For the present work, where the apparatus needed at times to be used completely out-of-doors, rectifier bridges were found more

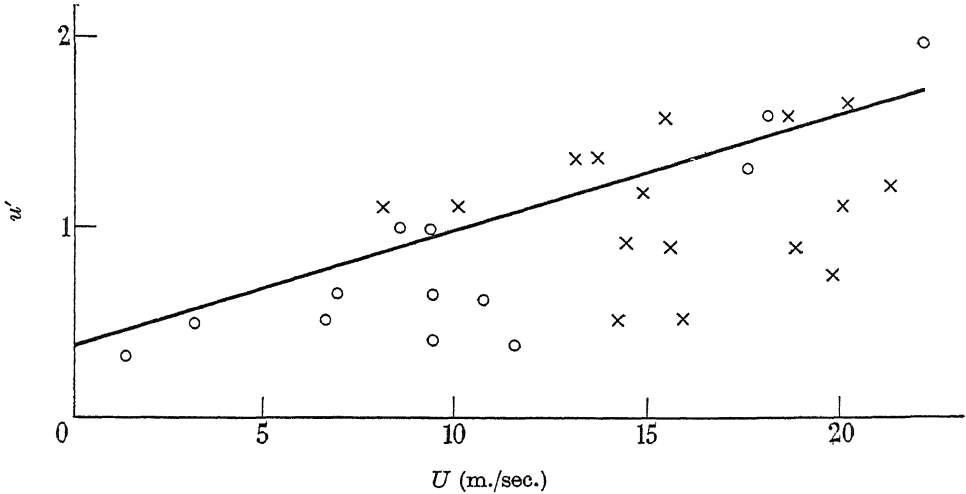


FIGURE 2. Overall gustiness: \times , Durham; O , Cairo.

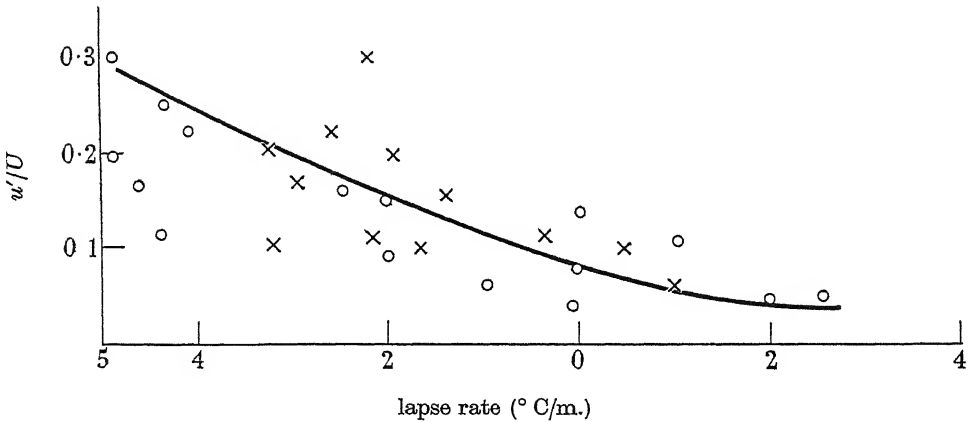


FIGURE 3. Gustiness and lapse rate: \times , Durham; O , Cairo.

suitable. In this circuit (figure 4) two copper oxide rectifiers are arranged with four wire resistances to form a bridge. The incoming alternating potential from the hot-wire u_1 is applied on one diagonal, and the output from the other diagonal goes to a voltmeter or to the amplifier and counter circuit. The second hot-wire is connected in the second bridge. To get a measure of the product $(u_1 u_2)$, the two bridges are connected in tandem as shown, with the meter across the combination.

This product meter was calibrated by applying to it, in place of u_1 and u_2 , the two output potentials from a Magslip to which a three-phase alternating potential of

constant amplitude about 1 V was applied. The power factor between the two outputs of the Magslip could be varied through known values by turning the rotor through a measured angle ϕ . It was thus found that, over a range of power factor between +1 and +0.2, a linear relationship existed between the log of the product

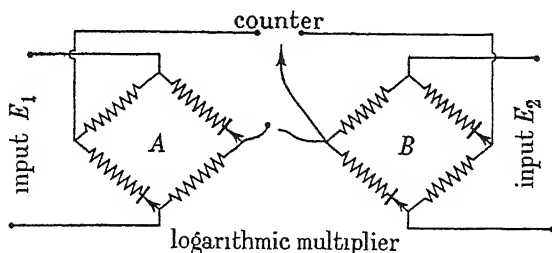


FIGURE 4. Logarithmic multiplier circuit.

$u_1 u_2$, i.e. $\cos \phi$, and the reading of a voltmeter connected as shown in figure 4. The apparatus is in fact a means of measuring the time integral of a product in decibels, and is so used in electronic computing machines. When, as in this case, the two wave forms to be multiplied are varying, the meter must have a natural period long compared to that of the waves in order to average the product:

$\cos \phi$	1	0.80	0.70	0.60	0.48	0.39	0.25	0.15	0.1
output of bridge	36	34	32	29	25	20	12	10	7

In applying this meter, the product $\overline{u_1 u_2}$ was measured on the tandem bridge and this result compared with the two separate outputs or with the measure of $\overline{u^2}$ obtained simultaneously on another bridge using an accessory hot-wire. In the light of the above calibration R was thus derived as a power factor.

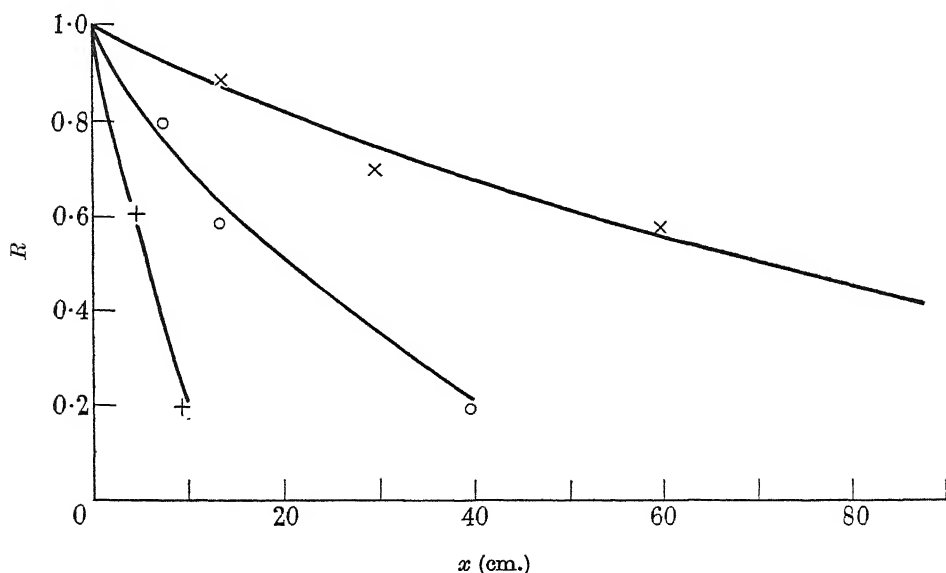


FIGURE 5. Correlation of velocity fluctuation with distance. Heights: x, 1500 cm.; O, 50 cm.; +, 10 cm. (not simultaneous records).

With this apparatus the variation of R with distance between stations at a horizontal level, both above the roof of the Observatory and nearer the grass plot, was traced at various wind speeds. Figure 5 shows specimen plots. From these it was possible to estimate the scale of turbulence defined as the length over which R falls nearly to zero. This length l may also be regarded as a measure of 'mean eddy size'. If the mean wind speed and the intensity of turbulence are also known, it is possible to calculate the eddy viscosity K from the formula $K = u' \int_0^\infty R dx$.

Further, Taylor (1922) has shown that in isotropic turbulence the mean rate of energy dissipation W is proportional to u'^2/l^2 ; thus:

TABLE 1

height (m.)	U	u'	l	K	W	lapse rate/m.
1500	250	25	150	3800	0.028	nil
50	150	15	60	900	0.06	-0.3
10	50	7	15	100	0.25	1

(Here the units of W are arbitrary, the remainder in c.g.s. units.)

The figures in the first row are a little surprising, considering the height above the ground, but were confirmed by the behaviour of the Dines anemometer at the time. This lies 2 m. above the hot-wire stations, though its 'effective height' by Air Ministry rating is considerably lower. The wires on this occasion were 3 m. above the Observatory roof and 3 m. back from its 'leading edge'. Owing to this 'site factor', the eddy lengths on the roof should not be compared with those over the grassy field.

Values of eddy viscosity can also be derived from measurements by Sheppard (1947) of the tractive force on a plate set in the ground, using the formula $\tau = K\rho dU/dy$. For example, at a time when the mean wind was 400 cm./sec., the lapse rate 1°C in 600 cm. and dU/dy unity, Sheppard found τ to be 0.2 dyne/cm.² leading to a value for K of 200 cm./sec.².

There was a characteristic difference between the mean eddy size near the ground than that at a height. On the roof of the Observatory eddies were usually longer than 3 ft., and even at a few feet above the ground, clear of obstacles, eddies were about 1 ft. in diameter, whereas closer to the ground than 1 ft., especially in the last 2 in., small eddies became much more common. Since, on the commonly accepted theory of the development of turbulence (Kolmogoroff 1941), large eddies transform into smaller ones, it appears that the eddies in the lower atmosphere have not the ground as origin. Only the smallest ones close to the ground evidently arise from surface friction.

Measurement of fluctuation components

It will appear in part II that in order to explain acoustic phase relationships over an area, knowledge of the vertical and horizontal components of turbulence is desirable. Two methods have been employed to measure v/w .

If the two components are to be measured by means of hot-wires, three wires meeting at an apex pointing upstream may be used, two meeting at about 15° in a horizontal plane to form an apex pointing up-wind and a similar pair in a vertical plane. Simmons & Bailey (1927) used such a pair in the two arms of a Wheatstone

bridge and showed that the bridge galvanometer registered direction, that is, v/U or w/U respectively. The relationship to galvanometer deflexion is not linear (cf. Simmons & Bailey 1927), but, as it is of the same general shape as the variation of potential across a single wire with mean wind, it can readily be linearized in the same type of valve amplifier. This was done in the present experiments, the two outputs being compared instant by instant on a double-beam oscillograph or in the X, Y display over an interval of time, as aforesaid.

The smoke apparatus is based on an instrument conceived by Frenkiel (1947). Smoke formed by bubbling dry air through titanium tetrachloride issues from a glass nozzle pointing down-wind into the atmosphere, where it forms smoke by hydrolysis to diffuse into the air. Two cameras, one vertically above pointing downwards and the other viewing horizontally to one side of the jet, allow the two aspects of the jet to be photographed over a period of, say, 15 sec., and the rates of spread are compared by photoelectric analysis of the two plates. It was found that inconsistent results were obtained with this apparatus, unless it was employed under isokinetic conditions, i.e. that the pressure in the bottle was adjusted within fairly precise limits to emit smoke into the air at the wind speed U .

TABLE 2

height above ground (cm.)	90	60	30	10	5
v/w	0.9	1.1	1	1.1	3

Both the smoke and the hot-wire apparatus indicate the preservation of nearly isotropic conditions in the turbulence of the air to within a foot of the ground, a conclusion which is somewhat at variance with the findings of Best (1935) that w is less than v even at 5 m. above the ground. It must be noted, however, that Best used a 'bidirectional vane' of considerable size, which could only respond to the grosser vagaries of the wind which constitute 'gustiness' in the meteorological sense and in which the horizontal vane, reading w , may have been damped in its motion by the proximity of the ground.

Although precise analysis of the frequency-spectra of the v and w records on a relative amplitude basis has not yet been possible, there is sufficient indication that the sphericity of the eddies scales with height on the basis of their diameter or that eddies have a ratio of vertical to horizontal axis which scales with the ratio of their mean diameter to height. The idea that the small eddies in turbulent flow are isotropic even when the large ones are not is another concept of Kolmogoroff's theory, confirmed here in the lower atmosphere.

Correlation of fluctuations of wind and temperature

It is evident that the spatial distribution of both air velocity and air temperature influence the propagation of sound. Especially does the latter affect the direction of sound rays in a vertical plane, since temperature has usually steeper gradients in the atmosphere in a vertical than in a horizontal direction. In fact, the mean value of the products vT^* and wT^* , where T^* represents the local deviation from the mean atmospheric temperature of the region, are significant for the refraction of sound waves in the two azimuths. The latter is the same factor which appears in the expression for the vertical flow of heat in turbulent diffusion, i.e. $q = \rho c_p \overline{wT^*}$,

ρ being the density and c_p the specific heat of the air. It has often been pointed out that T^* is not itself a measurable quantity, but the writer has found that one can measure, though not to the sensitivity with which one can measure v and w , the fluctuation of temperature at a point near the ground. It should therefore be possible to obtain a value for the temporal average of the product at any point, though, as far as the author knows, this has not heretofore been attempted. Corssin (1949) has discussed the theoretical possibilities of such a measurement.

Our method was to use a thermistor heated by a very low current (0.1 mA) for T^* and the usual hot-wire for u' . These were fed into the rectifier bridge in place of u_1 and u_2 to give $\overline{uT^*}$. In recording on the oscillograph it was found better to apply the T^* to a high-frequency a.c. bridge and to exhibit the resultant modulation of the supply on the screen. Thus figure 11, plate 2, shows simultaneous records of u (above) and T^* (below), and table 3 shows typical data obtained in this way.

TABLE 3

date, July 1949	...	19	23	26	28	
u'		5.5	24	7.5	5	30
$\overline{T^*}$		30	12	28	31	5.5
$\overline{uT^*}$		30	18	40	25	11
$R_{u, T}$		0.65	0.35	1	0.50	0.25
U		500	0	0	240	1200
lapse rate/m.		0	2	2	0	-1

(Here, the first three rows represent reciprocals of counter times, to correspond to those of the calibration table (see above p. 151), from which the values of R are derived. In the third example, R actually works out greater than unity, but it must be noted that the figures for u' , $\overline{T^*}$ and $\overline{uT^*}$ were obtained in quick succession and not simultaneously.)

At large lapse rates, U is small, T^* may be considerable and u is closely linked with it. Under these conditions marked divagations in time of sound rays passing nearly horizontally occur, leading to great differences in the simultaneous 'pick-up' of sounds at stations in a vertical plane (cf. part II). Small lapse rates or inversions occur mostly on occasions when both U and u are large and T^* is so small as to escape measurement and contributes little to the diffusion. Such conditions give fluctuations in sound intensity it is true, due to the magnitude of u , but they are more evenly distributed in space.

Since any resistance element reaches equilibrium at a resistance value dependent on ambient temperature and draught—in this case, on both T^* and u —it may be possible to heat it with such a current that it responds to some function of uT^* . One could then measure u , T^* and uT^* with a single element by raising the heating current successively.

An attempt to do this is illustrated in figure 12, plate 2, where the hot-wire (above) is heated by the usual current and the nearby thermistor (below) in its a.c. bridge is heated to 5°C above its surroundings. The latter follows the grosser variations of velocity that the former indicates, but has independent motions which are apparently due to temperature differences in the air passing by it. This technique is being further developed.

PART II. OPEN-AIR SOUND INTENSITY FLUCTUATIONS

The early investigators on the propagation of sound in the atmosphere, such as those engaged by the Trinity House Brethren for work on sirens or by the British army for work on gunfire, used either subjective observation or objective apparatus incapable of following rapid fluctuations. Consequently they were obliged to explain their results in terms of the continuing meteorological elements like mean wind and temperature, though they recognized that certain meteorological conditions led to an irregular signal when picked up at a distance of the sound from a steady source. It is only recently with the introduction of the sound-level meter and the cathode-ray oscillograph that studies of the fluctuations in the signal received from a steady source of sound have been possible.

Eagleson (1941), Sieg (1940) and a group working under Pielemeier (Schilling *et al.* 1947) have published reports within the period of the recent war in which attention was directed to such variations. Eagleson and Sieg worked with audible sources over distances ranging up to 500 m., noting at the same time the prevailing meteorological conditions such as temperature, wind and humidity. The other group worked in a jungle with ultrasonic sources up to 500 kc./sec., and made more detailed observations of air conditions. Krassilnikov (1945) attacked the same problem from a different point of view; he measured the difference in phase between a source and microphone and noted the fluctuations in phase difference. None of these workers attempted an instantaneous correlation between sound and meteorological records.

The general conclusions of the researches can be summarized as follows:

The magnitude and frequency of the fluctuations in the signal increase with length of track and frequency of the sound, amounting to as much as 50 % of the mean.

The fluctuations also increase with the 'roughness' of the terrain, being greater when the sound passes over standing crops than over a cropped plain, but in the jungle the signal is steadier, though more rapidly attenuated in the direct beam owing to scattering.

In the present research, the source was a loud-speaker supplied with a pure tone from a decade oscillator. The source was usually set at 4 ft. over the ground, though a limited number of observations were made with the source on a scaffold at 12 ft. The microphone was at the same height.

The source was usually directly down-wind of the microphones of the sound-level meters (Dawe type with filter-analyzers) at either 90 or 250 ft. Each microphone faced the loud-speaker in a housing lagged at its edges to prevent unwanted sounds originating at these points. Even so, sites and conditions in which the wind blew directly into the microphone proved difficult or impossible. Frequencies of 500, 1000, 2000, 4000 and 7500 c./sec. were used. The intensity near the source was about 90 db., except for the highest frequency at which the source gave only 50 db.

It was found essential to pass the received signal through a sharply tuned filter or analyzer so as to sort out the signal from the country background noise which was, of course, more serious at high frequencies. On these, indeed, work was only possible in high summer in the evening or early morning when insects were asleep. The output from the analyzer was shown either on a meter graduated in decibels or on a cathode-

ray oscillograph indoors, where it could be photographed. On the double-beam oscillograph these types of record were possible: (1) output from two meters at a distance from each other, exhibited one over the other or in X - Y relationship for phase; (2) output from one meter (above) and hot-wire midway down sound (below); (3) output from two hot-wires; (4) v and w averaged (cf. part I).

The object in this case was to compare the motional fluctuations in the local atmosphere with the acoustic ones. In a general way this was done by taking a reading on the turbulence meter, quickly followed by one in which the fluctuations in signal intensity were injected into the thyratron counter in place of the turbulence fluctuations. (This was possible because the output from one of the sound-level meters gave an envelope of intensity in the final stage, so that the potential across its output resistance could be fed to the counter in place of the potential across the hot-wire.)

Three types of terrain are involved: (1) at Durham over cropped grass; (2) at Durham over standing crops; (3) at Cairo over sand.

Typical records are shown. Figure 13, plate 2, are of velocity (above) and sound intensity (below) at 2000 (left) and 4000 c./sec. (right). Figure 15, plate 3, is similar at 4000 c./sec.—the sound in integrated form in another meter—to show the marked effect of the regular eddies in the wind (cf. figure 9, plate 2) on sound dispersion at this frequency. Figure 16, plate 3, shows the difference between sound of 4000 c./sec. picked up (left) well clear of the ground on a calm day, (right) close to the ground amid thermal currents.

For overall sound attenuation in the present research, the output from the microphones was fed into decibel meters with a slow response, so that a time average of intensity could be taken in line ahead of the source. Such measurements were taken with source and microphone at equal heights over the ground at levels varying from 1 to 12 ft. above the latter, which acted as a sound absorber.

Results at several frequencies for two types of terrain are exhibited in figure 6. Owing to the close intensity scale employed, necessitated by the desire to get all the results on one figure, the scatter of points is not well seen on these graphs. The difference in this respect between indoor and outdoor results can, however, be seen on a later graph (figure 8). The data of figure 6, obtained as they were in less gusty weather, had a scatter intermediate between those of the left and right of figure 8.

There are, of course, sources of loss which are independent of atmospheric turbulence or of absorption in the boundary layer over the ground. There is first the decay due to the spreading of the sound, which, on the assumption that spherical radiation persists after a certain distance, accounts for an attenuation of $20 \log r$ decibels for a distance r from the source.

There is an additional absorption, made clear by Knudsen (1935), when the air contains water vapour. This effect is a function of frequency, the 'molecular absorption' so occasioned rising to a maximum in the neighbourhood of a humidity of 0.2 g./l. and accounting for about 3 db. per 100 ft. at 1000 c./sec., waxing nearly in direct proportion to the frequency.

When 'fast' recording by the analyzers was tried, they were fed to the counter or to the cathode-ray oscillograph employing a slow-moving film so that the intensity

envelope over a minute or so could be recorded. At short intervals the loud-speaker was switched off so that 'background' at this frequency could also be recorded. At the same time, as already noted, a measurement of u' was made. The band width of

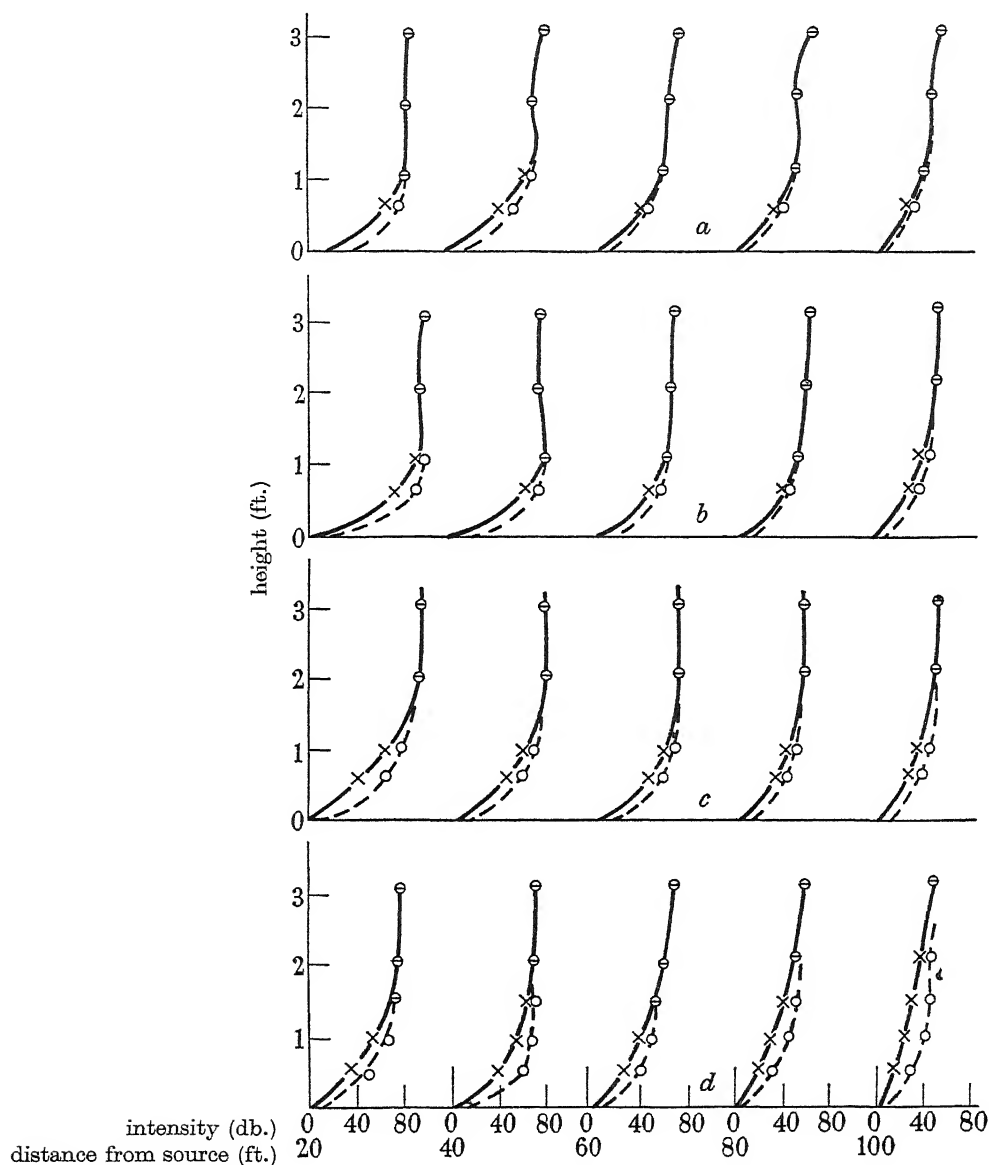


FIGURE 6. Absorption of sound with distance from source. —, over turnip crop; ---, over short grass. Frequencies (c./sec.): a, 1000; b, 2000; c, 4000; d, 7500.

this background intensity (i) was then correlated with u' on figure 7. The horizontal lines represent the spread of the points corresponding to a definite band width, with a curve drawn through the maximum frequencies of appearance of a given i in the data. An attempt to correlate sound fluctuation and lapse rate was less happy, but one general conclusion could be made. At high lapse rates the sound band width is

small, but the band is overlaid with microscopic variations which cause the band to waver over a considerable area in 1 min. Such a record is, in fact, characteristic of refraction rather than of dispersion. Under these circumstances, too, the background has small intensity compared to the main signal. On the other hand, the background is large and shows much indentation at small lapse rates or inversion. At zero lapse both band width and background intensity are large when u/U is large.

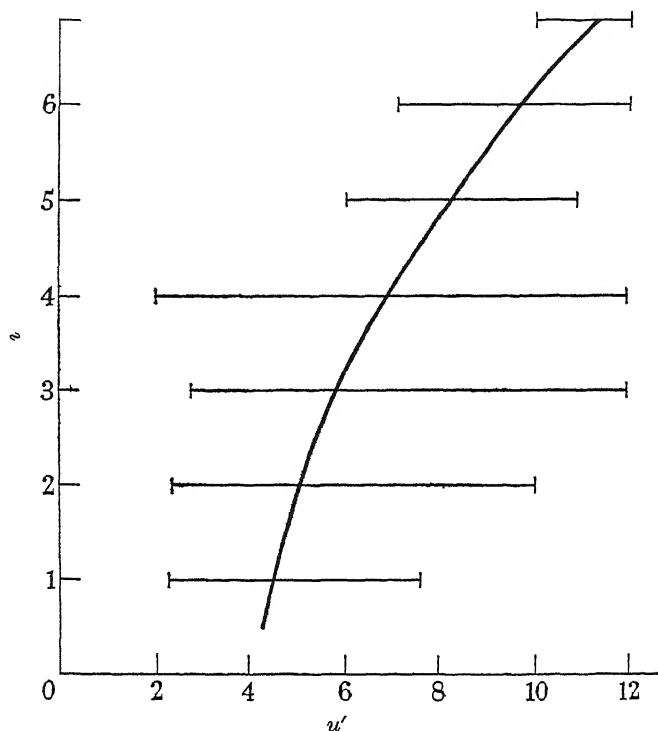


FIGURE 7. Background sound intensity at various wind speeds (arbitrary units).

There is a possibility of analyzing these background sounds, if we take care that they are entirely due to the wind-blown eddies, in order to obtain the spectrum of the turbulence. For this purpose, measurements of the 'country noise' on the Northumbrian moors have been made on a sound-level meter of 'flat response' in which the sounds picked up over the range 25 to 7500 c./sec. may be quickly recorded step-by-step. (It is, of course, necessary to ensure that no man-made or similar biological noises have sufficient intensity to be recorded. In summer on the moors one had to spray insecticide liberally in the vicinity of the noise meter.) Two typical spectra obtained in this way are exhibited in table 4.

It should be noted that this method gives an average effect over the whole region from which the natural noises can penetrate at sufficient intensity to affect the meter

TABLE 4

frequency (c./sec.)	25	50	100	200	500	1000	2000	5000
$U = 3$ m./sec., background I (db.)	25	20	20	15	15	10	5	5
$U = 12$ m./sec., background II (db.)	40	50	50	30	20	10	8	8

and cannot give readings at one height only. So far, however, these sound records of turbulence have been found to confirm in a general way the spectra derived from analysis of hot-wire records obtained at a moderate height from the ground.

Sound-phase fluctuations at two stations

Though the experiments so far described suffice to give a general picture of the relationship between sound propagation and turbulence in the atmosphere, they do not tell the whole story, inasmuch as our discussion has been limited to the intensity of the signal. Now an essential feature of sound, since it is propagated as wave motion, is the existence of phase relationships between the sounds arriving from a common source at different points in the field or at the same point by different routes, so that a steady source of sound inside a room in which the objects are stationary and the temperature is uniform sets up a steady field in which the intensities and phase relations between signals picked up at various points are invariable in time.

Our first experiments on the intensity at one location showed a different state of affairs in the open air, which was confirmed when two microphones and separate tuned analyzers had their outputs severally connected to the *X* and *Y* plates of the oscillograph. One microphone was placed straight ahead up-wind and at the same level as the source, while the other was raised to different heights (up to 30 ft. above the first) or at the same level to one side. Intensities and phases were compared over periods of 15 sec., and the records included some background which showed little or no correlation in phase.

The following general observations were made:

(1) The greater the fluctuations shown, under given conditions, on a single microphone, the more rapidly does phase correlation diminish with distance in either *y* or *z*, when two microphones are used.

(2) There is a close link between this reduction of phase correlation with distance and the intensity of turbulence, the deviations from 'classical' propagation being greatest under desert conditions, where the vagaries of the air in both temperature and the resulting agitation are a maximum.

(3) Especially under the same thermal conditions and especially near the ground, correlation falls off more rapidly in a vertical than in a horizontal line and, in either azimuth, more quickly over standing crops than over a bare field.

(4) A closer correlation exists at high frequencies. Figure 17, plate 3, which illustrates these last two points, refers to a calm day with a lapse rate of $2^{\circ}/\text{m.}$ near the ground. On another windier day, of zero lapse rate, all frequencies on two microphones at more than a foot separation showed the same type of record, that peculiar to the lowest frequency of the figure 17 records.

Under windy conditions the phase records were overlaid with much incoherent background proper to the frequency picked up, but one could still pick out the wanted signal as a more coherent figure, often slowly changing in form with a quasi-period of 10 to 30 sec., demonstrating a slow variation of path difference under the influence of the prevailing temperature and wind fluctuations.

The difference in the vertical and horizontal phase relationships close to the ground is paralleled by the difference in the ratio v/w with height noted in part I and

demonstrates the role of eddy diffusion in the propagation of sound out-of-doors. For instance, with source and microphone at 12 ft. horizontally, this anisotropy in signal propagation was considerably less, *ceteris paribus*, than when they were at a 6 ft. vertical separation.

The effects of refraction were also noticeable, commonly at Cairo, less often at Durham, when the mean intensity at two microphones in a vertical line were compared. When, in fact, the lapse rate was high, rays initially passing over the ground were bent up, coinciding at the upper microphone with others initially directed obliquely upwards, so that the upper microphone gained intensity at the expense of the lower.

In order to exemplify further the effect of diffusion on sound propagation, the field in planes across the front of a loud-speaker was explored in horizontal traverses with a small microphone, first in an anechoic chamber and again with the loud-speaker in the open over a grassy field. In the chamber (which was 15 by 25 by 12 ft. and lined with asbestos wool) the pattern conformed quite well to classical diffraction theory, except that the minima were not places of absolute silence, perhaps because of incomplete absorption at the walls and the finite size of the microphone, but in the open even on a calm day, the sound field departed considerably from this, the maxima and minima being smudged. On a gusty day one would have difficulty in recognizing any laws of diffraction; rather is the energy scattered by the eddies in the way that light is scattered by a suspension of transparent bodies of size comparable with the wave-length such as it encounters in a fog (cf. figure 8).

These departures from the classical laws of sound propagation are to be ascribed to the turbulent nature of the atmosphere. The normal velocity c of sound in still air or in steady streaming is locally replaced by $c + u$, consequently we should expect a scattering of the sound from the direct path on the basis of the mean correlation coefficient R of turbulent motions over the path from source to listening post and with amplitude proportional to u'/c . Attempts have been made to set up theories which regard the eddy length as playing the part of the regions of abnormal refractive index which appear in Rayleigh's theory of light scattering, with its dependence on l/λ and on the volume of the scattering entity in the form of l^3 or of $\int_0^\infty x^2 R dx$. There is, however, this difference, that the atmosphere is a medium of variable acoustic refrangibility without variable density—in the absence of thermal currents. This latter condition was not envisaged by Rayleigh, who, in fact, proceeds on the assumption that the normal component of particle velocity relative to the boundary of his scattering spheres must be made to vanish.

We should not therefore expect that the distribution of scattered sound round our 'secondary sources' will be the same as in optics, even when polarization is allowed for. It seems likely that more scattered, or rather, refracted, energy will pass into small angles relative to the direction of propagation. Nevertheless, these experiments show a general agreement with the optical analogy as far as dependence of the total deviated energy on l/λ and u'/c is concerned.

The main practical conclusion of this work is that there is a close relationship between the instantaneous patterns of air velocity and sound fields, but that the

true delineation of each requires continual records at several points. There is no short cut for predicting either from a few simple observations, for example, of lapse rate. The hope, first expressed by Professor O. G. Sutton, that one might predict the large-scale diffusion of matter in the atmosphere (of smoke, for example) by observations on the sounds picked up by microphones in the vicinity seems justified, yet the

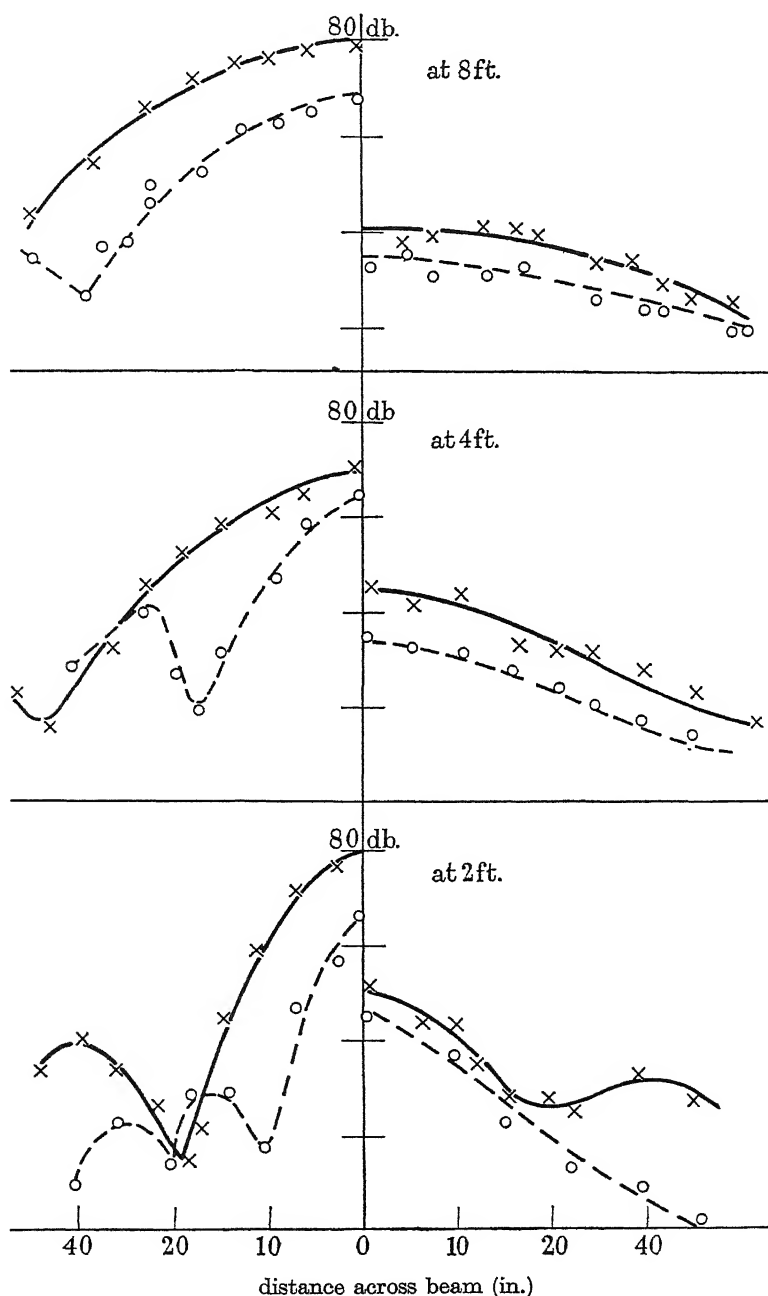


FIGURE 8. Diffraction of sound: in anechoic chamber (left); in open air (right).
—, 4000 c./sec.; ---, 7500 c./sec.

instrumentation of the two measurements, eddy viscosity and sound propagation, over an area of the order of a square mile, seems equally tedious. Both depend in a somewhat uncertain way on the *mean* meteorological elements for the area, and these will probably remain the data which the smoke engineers will have time to collect.

The author wishes to proffer his best thanks to the Research Fund Committee of King's College, Newcastle upon Tyne, for a grant to defray the expenses of this work, to the Senate of Fouad I University, Cairo, for a Visiting Professorship, which enabled him to start the desert observations, and to Assistant Professor M. Mokhtar and his students for continuing these after his return to England, finally, to Professor O. G. Sutton, F.R.S., for helpful comments on the first draft of the manuscript of part I.

REFERENCES

- Best, A. C. 1935 *Geophys. Mem.* 65.
 Carruthers, Nellie 1943 *Quart. J.R. Met. Soc.* 69, 289.
 Corssin, S. 1949 *Tech. Notes Nat. Adv. Comm. Aero., Wash.*, no. 1864.
 Eagleson, H. V. 1941 *J. Acoust. Soc. Amer.* 12, 427.
 Findeisen, H. 1936 *Met. Z.* 53, 6.
 Frenkiel, F. N. 1947 *C.R. Acad. Sci., Paris*, 224, 98.
 Godicke, K. 1935 *Ann. Hydrogr., Berl.*, 63, 400.
 Kolmogoroff, A. N. 1941 *C.R. Acad. Sci. U.R.S.S.* 30, 301.
 Knudsen, V. O. 1935 *J. Acoust. Soc. Amer.* 6, 119.
 Krassilnikov, V. A. 1945 *C.R. Acad. Sci. U.R.S.S.* 46, 103.
 Prandtl, L. & Reichardt, H. 1933 *Dtsch. Forsch.* 110.
 Schilling, H. K., Givens, M. P., Nyborg, W. L., Pielemeier, W. A. & Thorpe, H. A. 1947 *J. Acoust. Soc. Amer.* 19, 222.
 Serase, F. J. 1930 *Geophys. Mem.* 52.
 Sheppard, P. A. 1947 *Proc. Roy. Soc. A*, 188, 208.
 Sieg, H. 1940 *Elekt. Nachr.-Tech.* 17, 193.
 Simmons, L. F. G. & Bailey, A. 1927 *Phil. Mag.* 3, 81.
 Sutton, O. G. 1934 *Proc. Roy. Soc. A*, 146, 701.
 Taylor, G. I. 1922 *Proc. Lond. Math. Soc.* 2, 20 and 196.

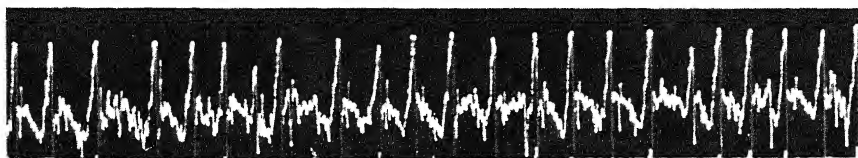
DESCRIPTION OF PLATES 2 AND 3

PLATE 2

- FIGURE 9. Fluctuation record showing presence of cellular vortices.
 FIGURE 10. Record showing turbulence 'burst' (above); sound of 7500 c./sec. (below)
 FIGURE 11. Velocity record near ground (above) and temperature (below).
 FIGURE 12. Hot-wire record (above) and thermistor record (below).
 FIGURE 13. Velocity and sound records: (left), 4000 c./sec.; (right), 7500 c./sec.

PLATE 3

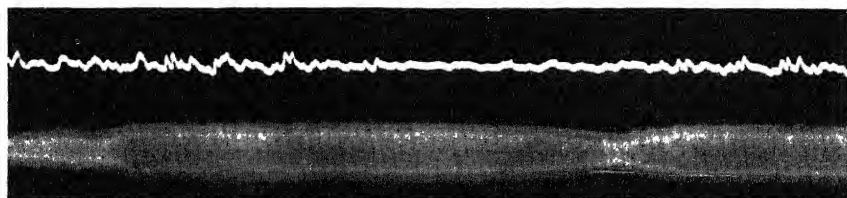
- FIGURE 14. Velocity fluctuations: on roof, $U = 2.1$ m./sec. (above); on ground, $U = 0.7$ m./sec. (below).
 FIGURE 15. Effect of cellular vortices (above) on sound of 4000 c./sec (below).
 FIGURE 16. Velocity and sound records at 4000 c./sec.: (left), near ground; (right), at height of 30 ft.
 FIGURE 17. Phase relation between two microphones: *a*, 1 ft. apart horizontally; *b*, 6 ft. apart horizontally, *c*, 4 ft. apart vertically; frequencies (left to right), 2000, 4000, 6000, 7500 c./sec.



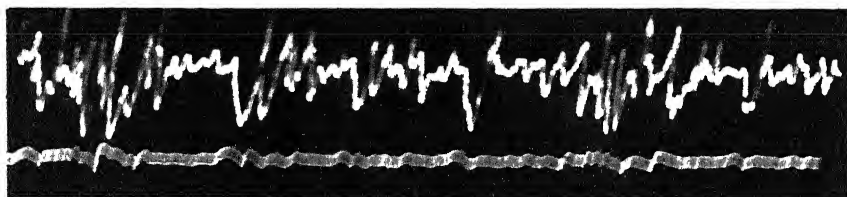
9



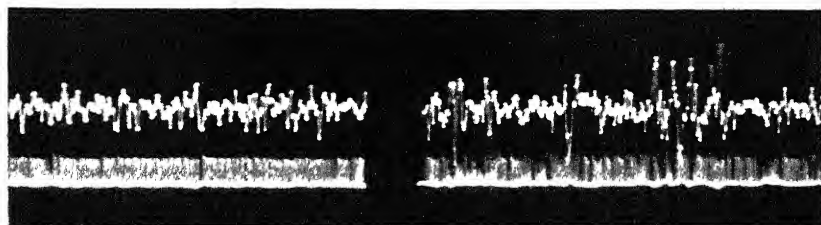
10



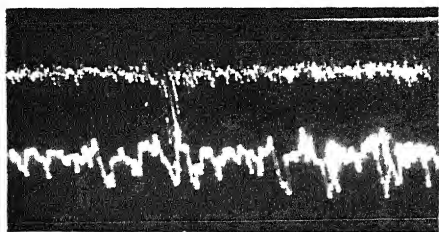
11



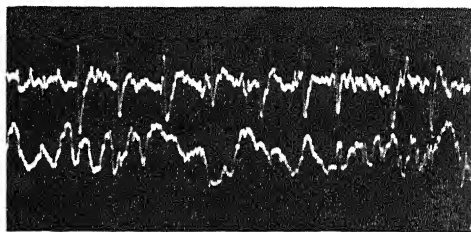
12



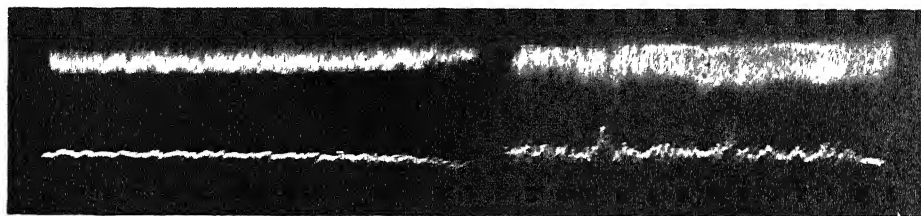
13



14



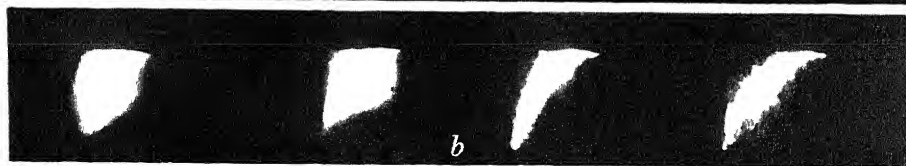
15



16



a



b



c

17

Diffusion and flow of gases and vapours through micropores

II. Surface flow

BY P. C. CARMAN AND P. LE R. MALHERBE

(Communicated by E. K. Rideal, F.R.S.—Received 25 January 1950)

Flow through porous plugs has been measured under conditions where surface flow of adsorbed gases makes an important contribution

An analysis of variables affecting surface flow enables surface diffusion coefficients to be calculated and has indicated conditions likely to reduce the correction for flow in the gaseous phase to small proportions. This has been confirmed by studying the effect of variations of porosity.

Surface diffusion coefficients tend to increase with 'coverage' of the adsorbent surface. This requires further study.

Activation energies measured for one system were found to be rather small. Owing to the variation of surface diffusion coefficients with coverage, it proved difficult to obtain accurate values. Activation energies indicated a tendency to decrease as coverage increases. Further experiments on these points are also being carried out.

For several years it has been generally recognized that molecules adsorbed at a solid surface are more or less free to move over the surface (see Brunauer 1943). Volmer & Adhikari (1925) demonstrated that the surface molecules of a benzophenone crystal are mobile, and, a little later (1926), that this applies also to benzophenone molecules adsorbed on glass. In the latter case, the resistance to diffusion was only about $\frac{1}{100}$ of that experienced by molecules of similar size in aqueous solutions. According to Bangham (1937), mobility is characteristic of physical adsorption, while adsorption on fixed sites is a criterion of chemisorption. As further indirect evidence of surface mobility, Bangham & Fakhoury (1930) has pointed out that swelling of a porous, rigid adsorbent such as charcoal, when physical adsorption takes place, can be accounted for quantitatively by assuming a thermal motion of adsorbed molecules, giving rise to a surface pressure.

Surface mobility must also play a part in kinetics of sorption by porous sorbents. If adsorbed molecules were fixed, the internal surface could only be reached by adsorbate molecules diffusing into the pore-space of the solid, followed by adsorption, and, in all cases of practical importance, such gaseous diffusion would soon become seriously hindered by molecules adsorbed at the outer parts of each adsorbent granule, since this would amount to constriction of the capillary passages leading to the interior. Simple calculations, assuming typical diffusion coefficients for gases, would indicate a very low practical efficiency for a bed of activated charcoal or silica gel granules, adsorbing from a stream or air, since, while equilibrium could readily be produced at the surface of each granule, the internal surface would be almost inaccessible. Experience is all to the contrary; for example, tests on adsorption of chloropicrin from air with charcoal in a military gas-mask Klotz (1946) show that the charcoal is nearly saturated before any noticeable concentration of chloropicrin passes through the bed. The most probable explanation is that molecules adsorbed at the surface undergo surface diffusion to the interior, and that this is a relatively

rapid process. The observation of Volmer & Adhikari on rates of diffusion of benzophenone on glass is in accord with this.

Surface diffusion implies a thermal motion of molecules adsorbed at a solid surface. It should be sharply distinguished from 'interstitial diffusion', encountered when a gas is sorbed by a medium in which pore diameters are comparable with the size of sorbed molecules, for example, by dehydrated chabazite (Barrer & Ibbitson 1944). It is not then possible to speak with certainty of 'adsorption' or of 'internal surface', the state of the sorbed molecules being more akin to solid solution than adsorption. Kinetics of interstitial sorption is governed primarily by the factor of molecular size. In fact, sorbents such as dehydrated chabazite act as 'molecular sieves', allowing only molecules below a certain size to pass. This molecular sieve action seems to have been produced by Rayleigh (1936) by flowing gases through the narrow passages, probably less than 10^{-7} cm., formed when two optically plane glass surfaces are pressed together. Comparing air and helium, the latter was found to flow 6.8 times as fast, presumably owing to its smaller molecular size. This is much larger than the ratio of 2.7 times to be expected from the relative molecular weights if Knudsen flow were taking place.

In contrast to both interstitial diffusion and Knudsen diffusion, surface diffusion tends to be most marked with larger and heavier molecules, since these are usually most easily condensed and adsorbed. Instances are given in the present paper where rates of flow of CF_2Cl_2 are much higher than for H_2 . In common with interstitial diffusion, it should be an activated process, though possibly with smaller energies of activation.

METHODS OF MEASUREMENT OF SURFACE DIFFUSION

Earlier studies by Damköhler (1935) and by Wicke (1939) were based on kinetics of sorption by granules of porous sorbents. Even assuming a simple diffusion coefficient, and a constant pressure of gaseous adsorbate, this is a fairly complex system, since conditions are continually changing within the pore-space of the sorbent. Since surface diffusion is always accompanied by normal diffusion and by Knudsen diffusion, the diffusion coefficient is not simple, and analysis of the experimental data is open to much uncertainty. In later papers Wicke (Wicke & Kallenbach 1941; Wicke & Voigt 1947) realized this and maintained a steady state of diffusion or of flow across plugs of porous sorbent, keeping conditions constant on each side of the plug. Flood & Tomlinson (1948) used the same method, and, in the present work, measurement was also made of the steady rate of flow through a porous plug, when the pressures, p_1 and p_2 , at the two free ends of the plug were kept constant. As will be explained shortly, this leads to a simple and direct calculation of coefficients of surface diffusion. To make this calculation, however, the adsorption isotherm must be determined and applied to obtain the difference of surface concentration across the plug, and two possible sources of error are therefore possible.

(i) Surface concentrations at the two free ends of the plug are obtained by assuming equilibrium at p_1 and p_2 . As adsorption and desorption are taking place, respectively, at a steady rate, equilibrium may not be maintained. In the present

work, flow-rates were deemed sufficiently low to ignore this possibility, but this assumption may not be justified.

(ii) Even though p_1 and p_2 are known accurately, *differences* of surface concentration depend very sensitively upon the accuracy of the adsorption isotherm, determined with a separate sample of the same powder.

Recently, an ingenious method has been devised by Haul (1950), in which a small drop of liquid, only about one-twentieth the quantity required to produce a monolayer, is placed on one end of the plug and its diffusion is followed by observing the variation of the centre of gravity with time on a micro-balance. In this way, surface concentrations are obtained without reference to the adsorption isotherm. The method can only be applied when the vapour pressure of adsorbate is negligible, i.e. under conditions which are completely different from those in the present paper. To this extent, the two methods are complementary. The new method is not free from difficulties. Calculation of surface diffusion coefficients from the observed data is not simple, and, if coefficients vary with coverage, e.g. as in table 2 of the present paper, calculations must be attended with uncertainty, since the essential feature of Haul's method is that coverage varies with time. Haul's discussion of his data does indeed indicate that this is a factor which cannot be ignored.

GENERAL EQUATION FOR FLOW IN MICROPORES

In part I (Carman 1950) it was pointed out that flow through micropores takes place by molecular streaming through the gas space. Only in table 9 of part I, using a very fine powder and easily condensed gases, CF_2Cl_2 and $\text{C}_2\text{F}_4\text{Cl}_2$, was measurable surface flow detected. Its effect is an enhanced rate of flow, i.e. surface flow takes place in parallel with flow in the gaseous phase. It is therefore desirable to establish unequivocally that the enhanced rate is indeed due to surface flow, and, if possible, to produce surface flow accompanied by little or no 'gas flow'. For this, an analysis of the variables affecting the different types of flow is essential.

We shall approach surface flow as an example of diffusion along a concentration gradient, obeying Fick's law. Assuming adsorption equilibrium at the end pressures, p_1 and p_2 , the quantities adsorbed are y_1 and y_2 millimoles/g. of adsorbent, where y_1 and y_2 are obtained from the adsorption isotherm for the temperature of the plug. As the weight of adsorbent per unit volume of plug is $\rho(1-\epsilon)$, concentration gradient in millimoles/cm.³/cm. = $\frac{\rho(1-\epsilon)\Delta y}{L}$, where $\Delta y = y_1 - y_2$. Then, if the flow-rate by surface diffusion is W millimoles/sec.

$$\frac{WL}{A} = K\rho(1-\epsilon)\Delta y, \quad (1)$$

where K is a diffusion coefficient with dimensions cm.²/sec.

The concentration gradient is arrived at by assuming equilibrium with the gaseous phase at the two free ends of the plug. We shall not inquire whether equilibrium is maintained between gaseous and adsorbed molecules in the interior, since it should not affect the issue.

If t sec. are required for the flow of 1 ml. at pressure p_1 and $T^\circ \text{K}$

$$W = \frac{1000p_1}{RTt},$$

whence, if $G = p_1/\Delta p t$,

$$W = \frac{1000\Delta p G}{RT}. \quad (2)$$

When there is no surface flow, permeability is governed by equation (3), which is equation (1) of part I when $k = 5$, $\delta/k' = 0.45$ and F_1 and F_2 are the porosity functions, $\epsilon^3/(1-\epsilon)^2$ and $\epsilon^2/(1-\epsilon)$:

$$\frac{GL}{A} = \frac{pF_1}{5\eta S_0^2} + \frac{0.96F_2}{S_0} \sqrt{\left(\frac{RT}{M}\right)}. \quad (3)$$

If now we assume that surface flow and gas flow take place independently and in parallel, the total permeability is obtained by substituting W for G as given by (2), and adding the surface flow term in (1)

$$\frac{WL}{\Delta p A} = K\rho(1-\epsilon)\frac{\Delta y}{\Delta p} + \frac{960F_2}{S_0\sqrt{MRT}} + \frac{200pF_1}{\eta S_0^2 RT}. \quad (4)$$

The last two terms are the slip-flow and viscous-flow terms, respectively, for flow in the gas phase. Together, they give the 'calculated gas permeability', i.e. the permeability if no surface flow were taking place. The viscous term is given last, since, in the present paper, powders are so fine that this term can nearly always be considered negligible.

The assumption of independent flow cannot be accurate, since gas flow must be blocked to some extent by adsorbed films. Equation (4), however, does indicate clearly how surface flow and gas flow are affected by different variables. It is thus possible (i) to decide whether differences between observed and calculated permeabilities are indeed due to surface flow, and (ii) to choose conditions under which the calculated permeability is a small enough fraction of the total permeability to be a negligible source of error. Consider the main variables in turn.

(a) Most important is the *specific surfaces*, S_0 . Increase of S_0 decreases gas permeability, but increases adsorption and hence $\Delta y/\Delta p$. Surface flow, in fact, can only make an important contribution when $S_0 > 10^6 \text{ cm.}^{-1}$

(b) At a given temperature, surface permeability depends on $\Delta y/\Delta p$, i.e. on the gradient of the adsorption isotherm, and hence on the *mean pressure*, p . Apart from the negligibly small viscous flow term, this variable does not affect gas permeability.

(c) A decrease of *porosity*, ϵ , rapidly decreases gas permeability, but tends to increase the surface permeability.

(d) A decrease of *temperature* has little effect on gas permeability, but tends to increase adsorption and hence surface permeability. On the other hand, it causes the surface diffusion coefficient to decrease.

EXPERIMENTAL

The apparatus described in part I was designed primarily for the present work. A similar apparatus was employed by Flood & Tomlinson (1948), but they used a charcoal rod. On comparing the surface by low temperature nitrogen adsorption,

1100 m.²/g., with the nitrogen permeability surface of only 64 m.²/g., it is clear that the permeability is much too high, owing to internal cracks and channels. Flow in the gaseous phase is thus greatly favoured and tends to mask surface flow. In the present work, using compressed plugs of fine powders, permeability surfaces and nitrogen adsorption surfaces were always at least of the same order.

The most outstanding requirement was long periods with constant pressures, p_1 and p_2 , in order to allow the surface concentration gradient to become steady, as shown by a constant flow-rate. The time for a single measurement varied with circumstances, but was never less than 3 hr. If the plug was initially clean, the apparent initial flow-rate could be more than ten times the final value; but, if it had adsorbed gas at a higher pressure previously, the apparent initial flow-rate was low or even negative, owing to desorption. In a few cases, the steady flow-rate was checked for reproducibility by approaching it from both sides.

Once a plug had been formed, it was not desirable to subject it to too drastic a treatment. To remove one gas, prior to introducing another, the apparatus was pumped out, and the permeability tube, if necessary, surrounded by a water-bath at about 50° C; but further attempts at degassing were not made, since, under conditions of flow, a plug can be regarded as self-cleaning. Of course, in the determination of adsorption isotherms, normal precautions were taken to obtain a clean surface.

Close control of temperature is essential, particularly when the saturation vapour pressure of the adsorbate is approached, e.g. under conditions employed in table 3, a change in pressure of 10 mm. or in temperature of 0.4° C would have produced an error of at least 20 % in the observed flow-rate. A large Dewar vessel was used as constant temperature bath, and four temperatures chosen, namely, water at 20° C, melting ice at 0° C, sodium chloride-ice eutectic at -21.4° C, magnesium chloride-ice eutectic at -33.1° C. The latter two were prepared and maintained by stirring finely crushed dry ice into solutions of eutectic composition, a method which enabled constant temperatures to be maintained, if necessary, for days on end, without removing the bath. No effect due to dissolved CO₂ could be detected.

Materials

Powders with a large surface and well-defined properties are desirable. Two powders were selected, namely, Linde silica and a carbon black, Carbolac 1.

Linde silica is an amorphous silica, density 2.2 g./ml., obtained by combustion of organic silicon compounds. The surface is normally about 300 m.²/g. by nitrogen adsorption, and the sample, Linde silica II, used in the present work, was in accord with this. Data in table 1, however, were obtained with a sample, Linde silica I, which had a surface of only 220 m.²/g. The electron microscope gives a particle size in agreement with the surface measurement and indicates a strong tendency for particles to aggregate in chains. Strong plugs covering a wide range of porosities are readily formed.

The sample of Carbolac 1 had a surface of 950 m.²/g. by nitrogen adsorption. Particles are not clearly resolved in the electron microscope, but appear to be 30 to 50 Å in diameter, which is in accord with the surface area. The density, determined by helium displacement, is 2.13 g./ml. Carbolac 1 is prepared by a hot air treatment

of a fine channel black, so that the surface is partly covered by chemi-sorbed oxygen, which accounts for about 10 % of the weight. Compared to Linde silica, the greater surface is an advantage, but plugs of Carbolac 1 were weak and had to be continually retested for development of internal cracks.

As absorbate, SO_2 was used in table 1, but subsequent experiments were carried out with Freon-12, CF_2Cl_2 .

Calculations

According to equation (4), the surface permeability is the difference between the total permeability, $WL/\Delta pA$, and the 'calculated gas permeability'. The latter is obtained by flowing a gas which is not adsorbed through the plug. Air could be used for Linde silica, but, as will be shown later, helium had to be used for Carbolac 1. If the viscous flow term is to be included, S_0 is calculated by equation (3), and S_0 is then used to calculate the 'slip term' and the 'viscous flow term' in equation (4). This was done in table 1, and the terms are given separately. In subsequent work, the viscous flow term could always be neglected, and the surface permeability was then calculated most simply by equation (5)

$$\text{surface permeability} = K\rho(1-\epsilon)\frac{\Delta y}{\Delta p} = \frac{WL}{\Delta pA} - \frac{W_1L}{\Delta p_1A} \sqrt{\left(\frac{M_1T_1}{MT}\right)}, \quad (5)$$

in which the last term is the 'calculated gas permeability', obtained from an experiment in which W_1 millimoles/sec. of an unadsorbed gas, molecular weight, M_1 , flow through the plug under a pressure difference of Δp_1 , and at a plug temperature of T_1 .

RESULTS AND DISCUSSION

Effect of pressure

Surface flow was first observed with SO_2 and a plug of Linde silica I, and results have been previously reported (Carman 1949). The data at 0°C are reproduced in table 1 in slightly different form. The table gives the observed total permeability, the slip term and the viscous term of the calculated gas permeability, and the surface permeability as the difference between observed and calculated values. Since p_1 was held constant, variations in the mean pressure were produced by varying p_2 . Corresponding variation in $\Delta y/\Delta p$ are given, and also the diffusion coefficient, K , calculated from equation (6):

$$K = \frac{\text{surface permeability } \Delta p}{\rho(1-\epsilon) \Delta y}. \quad (6)$$

The following points are illustrated.

(a) Surface permeabilities increase, as expected, with $\Delta y/\Delta p$. In particular, the rapid increase when the saturation pressure of SO_2 is approached on one side of the plug is to be noted.

(b) When p_2 exceeds the saturation pressure, condensation takes place on one side of the plug and produces an abrupt decrease in permeability. A similar result was found with CF_2Cl_2 at -21.4°C , and establishes clearly that condensation to liquid, in contrast to condensation to an adsorbed film, hinders diffusion.

(c) Diffusion coefficients of molecules of similar size and at similar temperatures in solution are of the order of 10^{-5} cm.²/sec., whereas K in table 1 has distinctly larger values. This would appear to support the conclusions of Volmer & Adhikari noted earlier.

 TABLE 1. FLOW OF SO₂ AT 0° C IN PLUG OF LINDE SILICA I

$\epsilon = 0.678$. S_0 by air permeability = 2.35×10^6 cm.⁻¹. S_0 by nitrogen adsorption = 4.48×10^6 cm.⁻¹.
 p_1 constant at 660 mm., p_2 varied. Saturation vapour pressure, $p_0 = 1161$ mm.

p_2 (mm.)	$WL/\Delta pA$ $10^{-10} \times$	calculated gas permeability		surface permeability $10^{-10} \times$	$\Delta y/\Delta p$ (p in mm.)	K (cm. ² /sec.) $10^{-4} \times$
		slip $10^{-10} \times$	viscous $10^{-10} \times$			
44	6.70	4.82	0.18	1.70	0.0030	1.08
90	6.82	4.82	0.20	1.80	0.0028	1.19
896	7.82	4.82	0.43	2.57	0.0048	1.00
935	8.30	4.82	0.43	3.05	0.0050	1.15
1067	9.54	4.82	0.47	4.25	0.0128	0.62
1160	18.9	4.82	0.50	13.6	—	—
1300	6.94	4.82	0.53	1.61	—	—

The main objection to the results in table 1 is that gas flow plays too large a part, so that surface permeabilities are too subject to error in the 'calculated gas permeability'. Later results, given in table 2, were obtained using Linde silica II, which has a larger surface, and compressing to a lower porosity, with the result that the lowest ratio of total to 'calculated' permeability is about 3. In this and subsequent experiments, it was preferred to keep Δp constant rather than p_1 when varying the mean pressure p .

 TABLE 2. FLOW OF CF₂Cl₂ AT -33.1° C IN PLUG OF LINDE SILICA II

$\epsilon = 0.497$. S_0 by air permeability = 5.4×10^6 cm.⁻¹. S_0 by nitrogen adsorption = 6.6×10^6 cm.⁻¹.
 Δp constant at 66 mm., p varies. Saturation vapour pressure, $p_0 = 660$ mm.

p (mm.)	$WL/\Delta pA$ $10^{-10} \times$	calculated gas permeability $10^{-10} \times$	surface permeability $10^{-10} \times$	Δy	K (cm. ² /sec.) 10^{-5}
41.9	1.58	0.57	1.01	0.53	1.52
169.0	1.79	0.57	1.22	0.34	2.85
323.0	2.26	0.57	1.69	0.38	3.54
421.4	3.24	0.57	2.67	0.62	3.41
520.4	9.09	0.57	8.52	1.20	5.64
547.3	15.3	0.57	14.7	1.29	9.05
573.9	25.5	0.57	25.0	1.43	13.9

The data in table 2 are plotted in figure 1, and an adsorption isotherm for the same temperature and a plug compressed to about the same porosity is given in figure 2. Values of Δy from this appear in table 2, and also values of K calculated by equation (6). Figure 1 confirms the extremely rapid increase in surface permeability as p approached p_0 , and this is *qualitatively* in accord with the increase in Δy . *Quantitatively* however, a number of unexpected features are shown.

(a) Surface permeability increased very much *faster* than $\Delta y/\Delta p$, shown by an increase of K of nearly nine times over the whole range.

(b) Only the values of K at high values of p are comparable in magnitude to those in table 1.

(c) Owing to the sigmoid shape of the isotherm, normal to multilayer adsorption, $\Delta y/\Delta p$ increases both towards low pressures and towards high pressures, but surface permeabilities continue to decrease down to the lowest observed values of p . Thus, at low pressures, the variation of surface permeability does not even agree in sign with the variation of $\Delta y/\Delta p$.

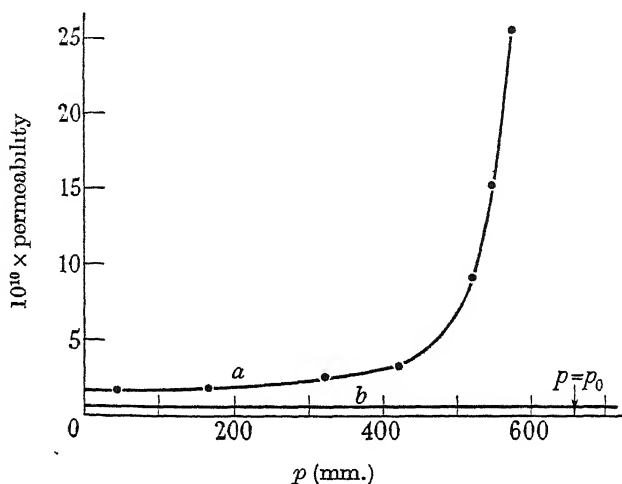


FIGURE 1. Permeability of Linde silica II, $\epsilon = 0.497$ towards CF_2Cl_2 at -33.1°C .
 a , total permeability; b , 'calculated gas permeability'.

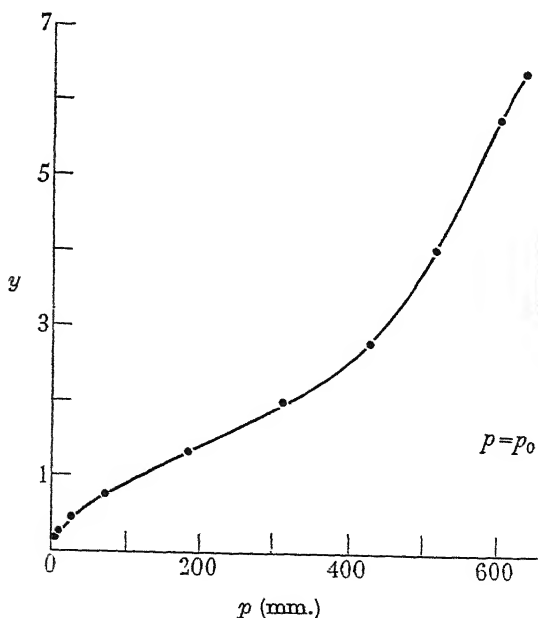


FIGURE 2. Adsorption of CF_2Cl_2 on Linde silica II plug. $\epsilon = 0.55$ at -33.1°C .

Table 1 gave no indication of similar behaviour, and it seems fairly certain that, in some cases at least, the relation between surface permeability and adsorption is more complex than postulated in equation (1). The most feasible explanation of the data in table 2 is that little or no flow takes place in the first layer of adsorbed molecules, and that flow in subsequent layers takes place with progressively greater ease.

Effect of porosity

To test equation (4) with respect to porosity, the temperature and pressures, p_1 and p_2 , were kept constant, so that variations in K could be excluded, and a series of plugs with different porosities was prepared. Data are summarized in table 3 and plotted in figure 3. Linde silica is particularly well suited to this purpose, as it gives strong plugs over an exceptionally wide range of porosities.

TABLE 3. FLOW OF CF_2Cl_2 AT -33.1°C IN PLUGS OF LINDE SILICA II AT VARIOUS POROSITIES

Δp constant at 66 mm. p constant at 547 mm. Δy constant at 1.29 millimoles/g.

porosity ϵ	L (cm.)	$WL/\Delta pA$ $10^{-10} \times$	calculated gas permeability $10^{-10} \times$	surface permeability $10^{-10} \times$
0.497	0.68	15.3	0.57	14.7
0.575	0.70	14.3	0.98	13.3
0.635	0.81	13.6	1.55	12.0
0.681	0.77	11.2	2.15	9.0
0.742	0.72	10.7	3.81	6.9
0.792	1.03	12.1	6.60	5.5
0.843	1.93	16.4	11.8	4.6

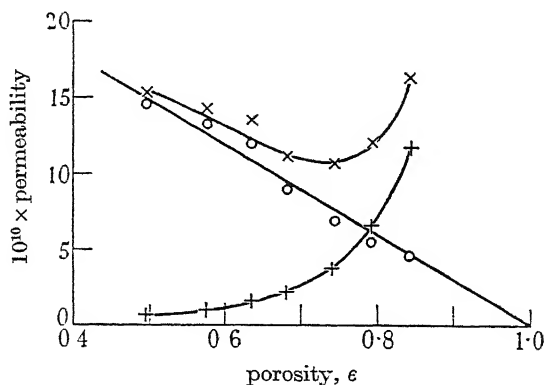


FIGURE 3. Effect of porosity on permeability of Linde silica plugs toward CF_2Cl_2 at -33.1°C . \times , total permeability; $+$, calculated gas permeability; \circ , surface permeability.

Calculated gas permeabilities were obtained from air permeabilities as shown in equation (5), and, in agreement with equation (4), decrease rapidly as ϵ decreases. Total permeabilities pass through a minimum and, when the difference, i.e., surface permeability, is plotted, the points lie reasonably close to a straight line passing through $\epsilon = 1.0$. Surface permeability thus increases as ϵ decreases, in proportion to $(1 - \epsilon)$, and this agreement with equation (1) affords the strongest evidence to date

that the phenomenon under consideration is indeed surface flow. In these experiments, plug lengths, L , varied over a considerable range, showing that, if Δp and Δy are both kept constant, K is not only independent of porosity, but also of the dimensions of the plug. Table 3 very clearly demonstrates the importance of low porosities in reducing gas flow to a small proportion of the total flow, and the lowest porosity is that of the plug used in Table 2.

Effect of surface

Since Carbolac 1 has about three times the surface of Linde silica, enhanced surface was expected, but the effects proved unexpectedly large. In table 4, the 'calculated gas permeability', $\frac{W_1 L}{\Delta p_1 A} \sqrt{\frac{M_1}{M}}$, of a tightly compressed Carbolac plug, with respect to CF_2Cl_2 at 20°C , is obtained from measurements with helium, hydrogen, nitrogen and carbon dioxide. In the case of Linde silica, all four gases gave the same values; but the results in table 4 can only be interpreted to mean that gases other than He are sufficiently adsorbed to produce a measurable rate of surface flow, even though H_2 and N_2 are far above their critical temperatures. For these two gases, permeabilities are independent of p . This is expected for Knudsen flow in the gas phase, and would also hold for surface flow, since Henry's law should apply under the experimental conditions, i.e. $\Delta y/\Delta p$ is independent of p . For CO_2 ,

TABLE 4. FLOW OF GASES IN CARBOLAC 1 PLUG AT 20°C , CALCULATED TO PERMEABILITY OF CF_2Cl_2 AT THE SAME TEMPERATURE

$\epsilon = 0.53$. Δp constant at 200 mm. S_0 by helium permeability $= 1.7 \times 10^7 \text{ cm.}^{-1}$.
 S_0 by nitrogen adsorption $= 2.0 \times 10^7 \text{ cm.}^{-1}$.

gas	p (mm.)	$\frac{W_1 L}{\Delta p_1 A} \sqrt{\frac{M_1}{M}}$ $10^{-11} \times$	K (cm. ² /sec.)
helium	126	1.96	—
	554	1.97	—
hydrogen	132	2.18	—
	558	2.19	—
nitrogen	124	3.36	—
	547	3.39	—
CO_2	127	7.74	—
	555	7.49	—
CF_2Cl_2	130	19.1	3.12×10^{-5}
	557	11.2	5.56×10^{-5}

an increase in p produces a decrease in permeability which exceeds experimental error, presumably because adsorption is large enough for the adsorption isotherm to be concave to the pressure axis. This is definitely the case for the permeability of CF_2Cl_2 , abstracted from table 5, for the same plug, since adsorption isotherms for this gas have been determined and are given in figure 4.

It will be noted that, in contrast with table 2, the variation of surface permeability at low values of the relative pressure, p/p_0 , is of the same sign as the variation of $\Delta y/\Delta p$. The agreement, however, is not quantitative, so that K does increase with

p , and is to this extent in agreement with table 2. Another point of concord is that, at low values of p/p_0 in each case, values of K are rather low. It would seem that, while there is a difference of mobility between successive adsorbed layers on both powders, it is much less marked for Carbolac, and, in particular, there must be a considerable mobility in the first layer, since the mobility of H_2 , N_2 and CO_2 could not otherwise be explained. From the experiments just cited, it is clear that the kind of surface as well as the extent of the surface has a profound effect on surface diffusion.

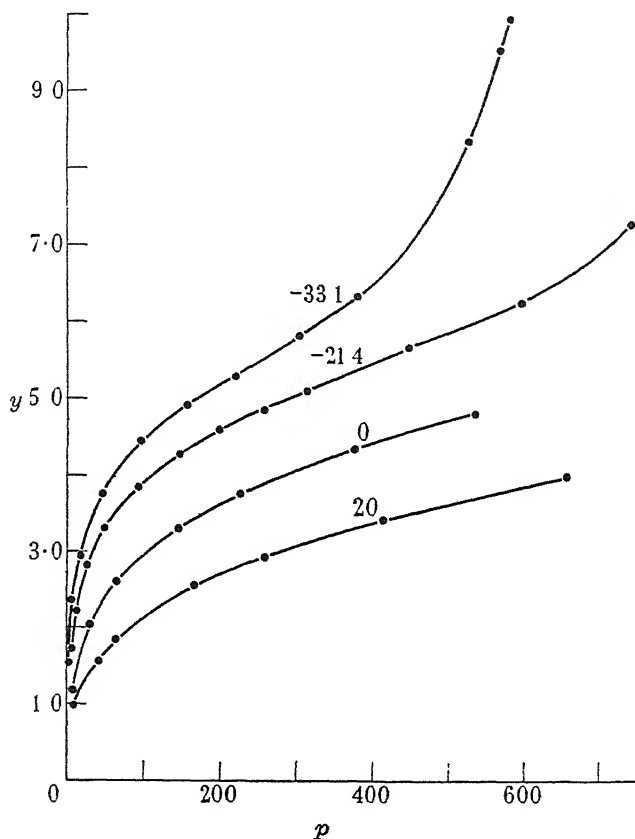


FIGURE 4. Adsorption of CF_2Cl_2 on Carbolac 1 at various temperatures ($^{\circ}C$).

Effect of temperature

Measurement of the temperature dependence of surface diffusion is very difficult, since the adsorption isotherm is itself so greatly affected by temperature. Bearing in mind the very great dependence of K upon p , i.e. upon the concentration of adsorbed gas, y , it seems clear that measurement must be compared at similar values of y_1 and y_2 at different temperatures. As the most important variable is the relative pressure, p/p_0 , it was hoped that it would be sufficient to compare permeabilities at constant values of p_1/p_0 and p_2/p_0 . Carbolac 1 was preferred since

(i) if p/p_0 is to remain constant over a wide range of temperatures, small values of p/p_0 must be chosen, in order to keep below the maximum pressure of 700 mm. for

which the apparatus is designed. With Linde silica, gas flow accounts for too large a proportion of the total flow-rate in this region, and uncertainties in allowing for it are correspondingly large.

(ii) Also, K for Linde silica is so sensitive to variations in y that temperature variations are likely to be obliterated.

Measurements in table 5 were therefore made on the same plug as in table 4. As plugs of Carbolac are very fragile, its permeability towards helium was re-checked after measurements at each temperature. Only one result is recorded at -33.1°C , since the plug was afterwards exposed to a high relative pressure at this temperature and was then found to have a high permeability towards helium.

The large amount of adsorption had presumably caused swelling and formation of internal cracks.

TABLE 5. FLOW OF CF_2Cl_2 AT VARIOUS TEMPERATURES, SAME PLUG AS IN TABLE 4.

temp. ($^\circ\text{C}$)	p_0 (mm.)	Δp (mm.)	p (mm.)	Δy	$\frac{1}{2}(y_1 + y_2)$	$WL/\Delta pA$ $10^{-10} \times$	Calculated gas permeability $10^{-10} \times$	Surface permeability $10^{-10} \times$	K $10^{-5} \times$	K corrected $10^{-5} \times$
$p/p_0 \approx 0.03$										
20	4250	208	129.6	1.52	2.21	1.91	0.20	1.71	3.12	3.42
0	2310	114	69.4	1.60	2.41	2.97	0.20	2.77	2.63	2.63
-21.4	1070	52.1	32.2	1.62	2.64	4.94	0.21	4.73	2.03	1.68
$p/p_0 \approx 0.13$										
20	4250	206	557	0.455	3.79	1.12	0.20	0.92	5.56	5.95
0	2310	114	301	0.42	4.07	1.58	0.20	1.38	4.98	4.98
-21.4	1070	52.2	137.1	0.37	4.21	2.58	0.21	2.37	4.45	4.24
-33.1	660	31.5	86.2	0.36	4.37	3.19	0.22	2.97	3.46	3.01

Two series of measurements were made, with $p/p_0 \approx 0.03$ and 0.13 , respectively. Both p and Δp were adjusted in proportion to p_0 to keep p_1/p_0 and p_2/p_0 constant. Adsorption isotherms plotted as y against p/p_0 , however, are not identical for different temperatures, so that y_1 and y_2 were not constant, as shown by mean values, $\frac{1}{2}(y_1 + y_2)$. Calculated gas permeabilities are those based on the helium permeability.

On plotting $\ln K$ against $\frac{1}{T}$, points were rather scattered, but the best average straight lines corresponded to energies of activation of 1520 cal./mole at $p/p_0 = 0.03$ and 1090 cal./mole at $p/p_0 = 0.13$. Since, however, K at a given temperature is not independent of p , i.e. of y , it is difficult to guess what errors arise from variations in y_1 and y_2 . A rough estimate is made in the last column, in which K is corrected to the value of y for 0°C , using the simple assumption that variation of K with y between $p/p_0 = 0.03$ and $p/p_0 = 0.13$ is linear. It can be seen that the correction is not negligible, and, in fact, temperature coefficients are so much changed that activation energies are increased to 2440 and 1680 cal., respectively. This uncertainty is not satisfactory and experiments are being undertaken to overcome this.

Isosteric heats of adsorption calculated from the isotherms in figure 4 showed a linear variation with y from 7900 cal. at $y = 1.80$ millimoles/g. to 5700 cal. at $y = 5.0$, and thereafter slowly approached the latent heat of condensation, 4840 cal.

Thus, surface diffusion has an energy of activation which is rather a small fraction of the heat of adsorption. This contrasts with the results of Haul (1950), who, for molecules of toluene, methanol and heptane on silica gel, found large energies of activation, of the order of 10,000 cal./mole. Wicke & Kallenbach (1941) have also found large energies of activation, but the approximations and assumptions used in their calculations were of necessity so crude that little weight can be given to them.

Effect of coverage

The variation of K with coverage shown in tables 2 and 5 was not expected, and clearly requires further investigation, particularly in view of the uncertainty produced by this factor in determining energies of activation. Experiments are being directed to a close study of the effect of coverage at a series of different temperatures. It should thus be possible to obtain activation energies at constant coverage with complete certainty, and to decide whether the variation of activation energy with coverage indicated by the results in table 5 is real.

Some relevant points may be made at this stage. The isotherm in figure 2 not only fits the general Brunauer-Emmett-Teller (1938) equation over the usual range, but it also fits the special form for n layers used by Pickett (1945) with remarkable fidelity over the whole range of p/p_0 , for values of n between 5 and 6. There seems no question that we are dealing with multilayer adsorption. The value of c in the BET equation is rather small, namely 11.5, so that there is considerable overlap between formation of the first layer and subsequent layers. A monolayer corresponds to 1.18 millimoles/g. and y reaches this value at p 154 mm. The variation of K with coverage indicates that layers are successively more mobile, but it is not possible to account quantitatively for changes in K by the simple assumption that the first layer is immobile and subsequent layers have constant mobility. It is only possible to say that there must be a large difference of mobility between first and subsequent layers, and that mobility of subsequent layers is not constant. Further analysis does not seem justified at the present stage.

Isotherms in figure 4 also obeyed the general BET equation over the normal range. Values of c were large in this case, indicating a high heat of adsorption for the first layer, as is borne out by the isosteric heats of adsorption given a little earlier. Compared with figure 2, therefore, overlap between formation of first and subsequent layers is relatively small. The various temperatures gave approximately the same capacity for a monolayer, namely, 3.9 millimoles/g. and y reaches this at values of p/p_0 from 0.082 at -33.1°C to 0.141 at 20°C . Comparing with values of p/p_0 in table 5, one reaches the conclusion that observed mobilities must have been almost entirely in the first layer and that K must therefore vary with coverage within this single layer. Another interesting observation is that values of c and therefore heats of adsorption of the first layer are very much less in the case of Linde silica than of Carbolac 1, yet mobility within this layer is more pronounced in the latter case. It would seem that the surface of Carbolac 1 is energetically much more uniform.

We wish to thank the South African Council for Scientific and Industrial Research for permission to publish this paper and also to acknowledge the gift of Carbolac 1 and

other carbon blacks from Godfrey L. Cabot, Inc., Boston, Mass. We also wish to acknowledge much useful help from Mr J. J. Theron in developing a sensitive and reliable relay system in the flow apparatus.

REFERENCES

- Bangham, D. H. & Fakhoury, W. 1930 *Proc. Roy. Soc. A*, **130**, 81.
 Bangham, D. H. 1937 *Trans. Faraday Soc.* **33**, 805.
 Barrer, R. M. & Ibbitson, D. A. 1944 *Trans. Faraday Soc.* **40**, 195, 206.
 Brunauer, S., Emmett, P. H. & Teller, E. 1938 *J. Amer. Chem. Soc.* **60**, 309.
 Brunauer, S. 1943 *Adsorption of gases and vapours*, **1**, 448.
 Carman, P. C. 1949 *Nature*, **163**, 684.
 Carman, P. C. 1950 *Proc. Roy. Soc. A*, **203**, 55.
 Damkohler, G. 1935 *Z. Phys. Chem. A*, **174**, 222.
 Flood, A. E. & Tomlinson, R. H. 1948 *Canad. J. Res. B*, **26**, 38.
 Haul, R. 1950 *Angew. Chem.* **62**, 10.
 Klotz, I. M. 1946 *Chem. Rev.* **39**, 241.
 Pickett, G. 1945 *J. Amer. Chem. Soc.* **67**, 1958.
 Rayleigh, Lord 1936 *Proc. Roy. Soc. A*, **156**, 350.
 Volmer, M. & Adhikari, G. 1925 *Z. Phys.* **35**, 170.
 Volmer, M. & Adhikari, G. 1926 *Z. Phys. Chem.* **119**, 46.
 Wicke, E. 1939 *Kolloidzshr.* **86**, 167.
 Wicke, E. & Kallenbach, R. 1941 *Kolloidzshr.* **97**, 135.
 Wicke, E. & Voigt, U. 1947 *Angew. Chem. B*, **19**, 94.

On the atomic theory of elasticity

BY KUN HUANG, *Department of Theoretical Physics, University of Liverpool*

(Communicated by M. Born, F.R.S.—Received 10 February 1950)

A scheme has recently been developed to present the general theories in lattice dynamics without specific assumptions about the atomic interactions. The present note aims at clarifying some basic points in using this scheme, as well as giving the correct expressions for the elastic constants (the available results given in previous works are shown to be correct only for central forces in spite of the use of the general scheme). The fact is emphasized that neither the potential energy of a homogeneous deformation from a reference configuration, nor the complete stresses in the configuration can in general be represented in the general scheme. Hence, for instance, the elastic constants cannot be deduced by straightforward means, nor the equilibrium conditions (vanishing stresses) imposed; the latter itself being necessary for the definition of the usual elastic constants.

A different technique is shown to be necessary for discussing such problems, one result obtained is that only five of the stresses in an arbitrarily chosen reference configuration can be explicitly represented, namely, all the anisotropic stresses. After introducing the condition that these stresses should vanish, the expressions for the elastic constants can be obtained, which are, however, to be used subject to the condition that the remaining stress, namely, an isotropic pressure, vanishes. Only when the general theory is applied to a concrete case, can the latter condition be explicitly introduced by the use of the given interaction.

The results are illustrated by the special example of central forces in the last section. The Cauchy relations follow as an incidental result; the assumptions upon which the relations rest are clearly exhibited in the simple derivation.

1. INTRODUCTION

It has been usual in standard treatments of the lattice theory of crystals (cf. Voigt 1928; Born 1923; Born & Göppert-Mayer 1933) to assume some specific types of atomic interaction, such as central forces, or interactions between oriented molecules, etc. In view of the great diversity of binding forces which are known to occur in different types of solids, clearly no such assumptions will prove generally valid. Much that is basic in the lattice theory is, however, common to all types of crystalline solids independent of the particular nature of the binding forces. Thus in more recent developments of the theory, attempts have been made, particularly by Born and his collaborators, to use, as the effective potential energy of a lattice, a general function Φ of the nuclear displacements (see, for example, Born 1942, 1945; Begbie & Born 1947; Smith 1948, Seitz 1940). Upon the function no further assumptions are made apart from the invariance requirements for displacements (translation and rotation) of the lattice as a whole, and the basic assumption of periodicity. Since the actual condition near the surface cannot affect discussions of the bulk properties, one could use the idealized model of an infinitely extended lattice. The nature of the potential is specified by giving its various derivatives

$$\left(\frac{\partial \Phi}{\partial u_{\alpha} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)} \right)_0, \quad \left(\frac{\partial^2 \Phi}{\partial u_{\alpha} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) \partial u_{\beta} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)} \right)_0, \quad \dots,$$

with respect to all the nuclear displacements $u_{\alpha} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)$, for k th particle in the l th cell

denoted as particle $\left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)$, for a reference configuration. The reference configuration is usually assumed to correspond to equilibrium. It is peculiar to an infinite lattice model that the equilibrium condition is twofold:

- (i) every lattice point is in equilibrium;
- (ii) all stresses vanish.

The condition (ii) follows from (i) in a finite lattice; but in an infinite lattice model, (ii) constitutes a separate condition (see example in §2).

The necessity and significance of imposing (ii) have been clearly described before (Born 1923, 1945),* but its consequence has not been correctly followed up in the general theory. The condition (i) is obviously equivalent to requiring

$$\left(\frac{\partial \Phi}{\partial u_{\alpha} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)} \right)_0 = 0. \tag{1.1}$$

From a configuration already satisfying (1.1), let us consider a homogeneous deformation

$$u_{\alpha} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) = \sum_{\beta} u_{\alpha\beta} x_{\beta} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right), \tag{1.2}$$

* In particular, in the theory assuming central interactions, Born has long ago (Born 1923) not only emphasized the condition (ii) but has given it also in explicit form in terms of the interactions. The following considerations refer thus to the situation one encounters in trying to achieve the same in the general theory, where only derivatives of Φ are assumed known.

where $\mathbf{x} \binom{l}{k}$ is the position vector of particle $\binom{l}{k}$ before the deformation. If one *could* express the energy density due to the deformation as a function of $u_{\alpha\beta}$, then in order to satisfy (ii) one had only to demand that the first derivatives of the deformation energy with respect to $u_{\alpha\beta}$ should vanish. Hitherto it has been assumed that this can be done, namely, by expanding Φ in a Taylor series in the displacements (1.2) with the help of the various derivatives of Φ . However, it is easily seen that as regards (ii) this procedure leads to a null result. For, once (1.1) is fulfilled, there will be no linear terms of $u_{\alpha\beta}$ in the formal expansion. This would imply that (ii) follows from the fulfilment of (i), but this is directly contradicted by the simple example given in §2. The fault in this procedure is quite obvious: the formal expansion only gives the energy change of the infinite lattice, which is, of course, a divergent quantity; in order to obtain a reliable result, one has to normalize it to a finite volume (cf. Born 1923). It is easy to convince oneself that one cannot obtain the desired normalized energy from the formal expansion; the difficulty is due to the ever-increasing displacement (1.2) as one proceeds farther away from the origin.

By way of a simple example it will be shown in §2 that in general *it is actually impossible to represent the deformation energy in the general theory*. Hence one cannot by straightforward means either impose the condition (ii) or deduce the elastic constants (second derivatives of energy density with respect to $u_{\alpha\beta}$).

The expressions for the elastic constants have been given before by Begbie & Born (1947, quoted as (I)) for non-ionic crystals and by Huang (1949, quoted as (II)) for ionic crystals. The methods used are based on the comparison of the ordinary elastic wave equations (with the consideration of piezoelectric complications in the ionic case) with the equations for the limit of long acoustic lattice waves deduced from the general lattice theory. The validity of these results is, however, restricted probably only to central interactions, because a symmetry relation used was based on an identity obtained by using arguments involving divergent expressions in the manner sketched above. The symmetry relation is found to be satisfied in the case of central interactions, but only after explicitly using the condition (ii). Since in the derivation of the identity, (ii) was not used, the result is obviously not reliable. (It is perhaps well to emphasize that we are referring here to the failure thus far of attempts to give the *general* expressions for the elastic constants, which hold for arbitrary interactions. Various existing calculations of elastic constants for specific types of crystals are of course not subject to the above difficulty, because certain assumptions concerning the forces are made to begin with, so that the energy *density* of a homogeneously deformed specimen can be obtained by special considerations.)

It will be shown that these various difficulties can be overcome by correctly developing the long-wave method. One finds, for instance, that without the symmetry relation we have mentioned, the lattice wave equation and the ordinary elastic wave equation are not in general compatible. In §3 the compatibility conditions are deduced which appear in the form of certain symmetry relations. Since the ordinary elastic wave equation presumes that there are no existing stresses (see §4), these compatibility relations express obviously *in part* or *in full* the condition (ii). When one assumes these relations fulfilled, the explicit expressions for the

elastic constants are uniquely determined by the comparison. Hence these are the correct expressions, subject of course to the condition that the configuration has actually no stresses.

In order to decide whether or not the compatibility conditions express condition (ii) *completely*, in §4 the lattice wave equation is compared to the elastic wave equation in a medium under an arbitrary system of homogeneous stresses. By this means it is shown that the five anisotropic stresses can be explicitly given in the general theory, and the compatibility relations actually express the conditions that these stresses should vanish. *The homogeneous pressure is not expressible in the general theory.* The condition for the latter to vanish can only be introduced explicitly when the general theory is applied to an actual case. From the comparison are also obtained ten identical relations for the second derivatives of Φ , expressing the invariance of the deformation energy against rotations.

In §5 the results are applied to the special example of central interactions which is thought advisable owing to the indirect procedure by which the results have been established.

2. THE EXAMPLE OF A LINEAR CHAIN

By way of a simple example, we shall illustrate the peculiar difficulties in the general theory. Consider a chain of identical particles interacting with the central potential $\phi(r)$ and take as a reference configuration with the particles at s apart along a straight line. The particles are numbered from an arbitrarily chosen particle as $l = -\infty \dots -1 \ 0 \ 1 \ 2 \dots \infty$. The displacements are restricted along the chain and denoted by u_l .

A homogeneous deformation is obviously described by

$$u_l = l\delta, \quad (2.1)$$

where δ is the increment in the next-neighbour distance. The energy change per particle can obviously be obtained from the interaction energy of particle 0 with all others as

$$\Delta E = \frac{1}{2} \sum_{l \neq 0} \phi(|ls + u_l|) - \phi(|ls|), \quad (2.2)$$

where always the absolute value is used in the argument. Making use of the fact that

$$\begin{aligned} \frac{\partial |ls + u_l|}{\partial u_l} &= 1 \quad (l > 0), \\ &= -1 \quad (l < 0), \end{aligned} \quad (2.3)$$

one can express (2.2) readily as a series in the displacements

$$\Delta E = \frac{1}{2} \delta \sum_{l \neq 0} |l| \phi'(|ls|) + \frac{1}{4} \delta^2 \sum_{l \neq 0} l^2 \phi''(|ls|) + \dots \quad (2.4)$$

The tension in the chain can be obtained as the energy change per unit increment in δ , thus

$$T = \left(\frac{\partial \Delta E}{\partial \delta} \right)_0 = \frac{1}{2} \sum_{l \neq 0} |l| \phi'(|ls|). \quad (2.5)$$

If one works with the scheme of the general theory, *the only quantities supposed known are the derivatives of the total potential function Φ with respect to the nuclear*

displacements. The derivatives can be obtained by differentiating the *formal* (divergent) expression of Φ :

$$\Phi = \frac{1}{2} \sum_{l'} \sum_{l \neq l'} \phi(|(l' - l)s + u_{l'} - u_l|). \quad (2.6)$$

One obtains easily, using (2.3), that

$$\begin{aligned} \left(\frac{\partial \Phi}{\partial u_l} \right)_0 &= \sum_{l' \neq l} \left(\frac{\partial \phi(|(l' - l)s + u_{l'} - u_l|)}{\partial u_l} \right)_0 \\ &= - \sum_{l' > l} \phi'(|l' - l|s) + \sum_{l' < l} \phi'(|l' - l|s) = 0, \end{aligned} \quad (2.7)$$

which is in fact obvious from the symmetry. For the higher derivatives one has for instance

$$\begin{aligned} \left(\frac{\partial^2 \Phi}{\partial u_l \partial u_{l'}} \right)_0 &= -\phi''(|l' - l|s) \quad (l' \neq l), \\ &= \sum_{l'' \neq l} \phi''(|l'' - l|s) \quad (l' = l), \\ &\text{etc.} \end{aligned} \quad (2.8)$$

In general, in the expression for a derivative of Φ of the n th order, only the n th derivative of ϕ appears.

In this example, there are two points to note:

(a) For the reference configuration chosen, namely, all particles equally apart, (2.7) shows that every particle is at equilibrium so that the condition (i) described in the introduction is already fulfilled. Nevertheless, the tension given by (2.5) can have any value. Hence, in general, for an infinite lattice model the requirement of vanishing stresses clearly constitutes an independent condition.

(b) Since $(\partial \Phi / \partial u_l)_0$ identically vanishes, the quantities given in the general scheme no longer involve the first derivative of ϕ . But (2.4), (2.5) show that both the energy of deformation and the tension in the reference configuration depend on ϕ' . It is thus clear that, in general, these quantities are not completely expressible in the general theory. This latter fact, as we have already pointed out, makes it impossible by straightforward means either to impose the equilibrium condition (ii) (cf. §1) or to deduce the elastic constants. It will be shown in the next two sections how these can be done with the help of the equations for long acoustic waves.

3. COMPARISON WITH ELASTIC WAVES IN MEDIA FREE OF STRESSES. COMPATIBILITY CONDITIONS AND THE ELASTIC CONSTANTS

In media free of stresses, there are twenty-one independent elastic constants (for media under stresses see §4) which may be expressed either in tensor notations as $c_{\alpha\gamma, \beta\lambda}$ ($\alpha, \gamma, \beta, \lambda = 1, 2, 3$) which satisfy the symmetry relations

$$c_{\alpha\gamma, \beta\lambda} = c_{\gamma\alpha, \beta\lambda} = c_{\beta\lambda, \alpha\gamma}, \quad (3.1)$$

or in Voigt notations as $c_{\rho\sigma}$ ($\rho, \sigma = 1, \dots, 6$) which satisfy the symmetry relations

$$c_{\rho\sigma} = c_{\sigma\rho}. \quad (3.2)$$

$c_{\alpha\gamma, \beta\lambda}$ are simply related to $c_{\rho\sigma}$ by

$$c_{\alpha\gamma, \beta\lambda} = c_{\rho\sigma} \quad (3.3)$$

with the following correspondence between the two sets of indices:

$$\left. \begin{array}{cccccc} \alpha\gamma \text{ (or } \beta\lambda) & 11 & 22 & 33 & 23(32) & 31(13) & 12(21) \\ \rho \text{ (or } \sigma) & 1 & 2 & 3 & 4 & 5 & 6 \end{array} \right\} \quad (3.4)$$

The elastic wave equation for a plane wave with the amplitude \mathbf{u} and wave-number vector \mathbf{y} can be expressed in tensor notations as (cf. II):

$$\rho \omega^2(y) u_\alpha = 4\pi^2 \sum_{\beta} \left(\sum_{\gamma\lambda} c_{\alpha\gamma, \beta\lambda} y_\gamma y_\lambda \right) u_\beta, \quad (3.5)$$

where $\omega(y)$ is the frequency and ρ the density.

Equations have been deduced for long acoustic lattice waves by Begbie & Born (I) for non-ionic crystals and by Huang (II) for ionic crystals, using the general theory. If the piezoelectric field is ignored in the equation given by Huang, these equations can be written in both cases as

$$\rho \omega^2(y) u_\alpha = 4\pi^2 \sum_{\beta} \left(\sum_{\gamma\lambda} \{\alpha\gamma, \beta\lambda\} y_\gamma y_\lambda \right) u_\beta. \quad (3.6)$$

The curly bracket expressions are certain combinations of the second derivatives of the formal potential Φ of the infinite lattice. In order to exhibit its full symmetry, we shall here give it in the following form:

$$\{\alpha\gamma, \beta\lambda\} = [\alpha\beta, \gamma\lambda] + (\alpha\gamma, \beta\lambda). \quad (3.7)$$

Their explicit expressions are

$$\left. \begin{aligned} [\alpha\beta, \gamma\lambda] &= -\frac{1}{2v_a} \sum_l \sum_{kk'} \Phi_{\alpha\beta} \left(\frac{l}{kk'} \right) x_\gamma \left(\frac{l}{kk'} \right) x_\lambda \left(\frac{l}{kk'} \right), \\ (\alpha\gamma, \beta\lambda) &= \frac{1}{v_a} \sum_l \sum_{kk'} \left\{ \sum_{l'} \sum_{k_1 k_2} \sum_{\mu\nu} \Phi_{\alpha\mu} \left(\frac{l}{kk_1} \right) x_\gamma \left(\frac{l}{kk_1} \right) \Gamma_{\mu\nu}(k_1 k_2) \Phi_{\nu\beta} \left(\frac{l'}{k_2 k'} \right) x_\lambda \left(\frac{l'}{k_2 k'} \right) \right\}, \end{aligned} \right\} \begin{array}{l} \text{non-ionic} \\ \text{case;} \end{array} \quad (3.8) \quad (3.9)$$

$$\left. \begin{aligned} \alpha\beta, \gamma\lambda] &= \frac{1}{8\pi^2 v_a} \sum_{kk'} \bar{C}_{\alpha\beta, \gamma\lambda}^{(2)}(kk') \sqrt{(m_k m_{k'})}, \\ \alpha\gamma, \beta\lambda) &= \frac{1}{4\pi^2 v_a} \sum_{k'' k'''} \sum_{\mu\nu} \Gamma_{\mu\nu}(k'' k''') \left(\sum_k \bar{C}_{\mu\alpha, \gamma}^{(1)}(k'' k) \sqrt{m_k} \right) \left(\sum_{k'} \bar{C}_{\nu\beta, \lambda}^{(1)}(k''' k') \sqrt{m_{k'}} \right), \end{aligned} \right\} \begin{array}{l} \text{ionic case.} \end{array} \quad (3.10) \quad (3.11)$$

The symbols in (3.8) and (3.9) will be explained in §5, when the results are applied to central interactions; to save space the reader is referred to the original paper (II) for the definition of the symbols in (3.10) and (3.11). For the moment it suffices to know the following symmetry relations between the new bracket expressions which we have introduced:

$$[\alpha\beta, \gamma\lambda] = [\beta\alpha, \gamma\lambda] = [\alpha\beta, \lambda\gamma], \quad (3.12)$$

$$(\alpha\gamma, \beta\lambda) = (\gamma\alpha, \beta\lambda) = (\beta\lambda, \alpha\gamma), \quad (3.13)$$

i.e. both brackets are symmetric in both pairs of indices, whereas the round brackets have the additional symmetry for an interchange between the pairs. Equation

(3.12) follows directly from the definition of the quantities involved and (3.13) is the consequence of the invariance condition for rigid rotations.

The lattice wave equation (3.6) has been obtained only under the assumption that all the first derivatives of Φ vanish (equilibrium condition (i) in the Introduction). The elastic wave equation (3.5), together with the symmetry requirements (3.3) on the elastic constants, holds, on the other hand, only for a medium free of stresses; hence let us *assume* that the reference configuration underlying the lattice wave equation also fulfils the condition of vanishing stresses (equilibrium condition (ii)). *Under this assumption*, equation (3.6) must be completely equivalent to the equation (3.5) together, of course, with the respective symmetry requirements on the coefficients appearing in the equations. Comparing the coefficients in these equations, one obtains

$$\sum_{\gamma\lambda} c_{\alpha\gamma, \beta\lambda} y_\gamma y_\lambda = \sum_{\gamma\lambda} \{\alpha\gamma, \beta\lambda\} y_\gamma y_\lambda, \quad (3.14)$$

or, since y is arbitrary,

$$\begin{aligned} c_{\alpha\gamma, \beta\lambda} + c_{\alpha\lambda, \beta\gamma} &= \{\alpha\gamma, \beta\lambda\} + \{\alpha\lambda, \beta\gamma\} \\ &= 2[\alpha\beta, \gamma\lambda] + (\alpha\gamma, \beta\lambda) + (\alpha\lambda, \beta\gamma). \end{aligned} \quad (3.15)$$

These relations must be satisfied subject to the symmetry relations (3.1), (3.12) and (3.13) for the elastic constants and the bracket expressions.

Let us regard (3.15) as equations determining the elastic constants $c_{\alpha\gamma, \beta\lambda}$. Since $(\alpha\gamma, \beta\lambda)$ satisfy the same symmetry relations as $c_{\alpha\gamma, \beta\lambda}$, let us write

$$c_{\alpha\gamma, \beta\lambda} = d_{\alpha\gamma, \beta\lambda} + (\alpha\gamma, \beta\lambda) \quad (3.16)$$

and substitute (3.16) in the equations (3.15). Thus $d_{\alpha\gamma, \beta\lambda}$ must be solutions of the following equation:

$$d_{\alpha\gamma, \beta\lambda} + d_{\alpha\lambda, \beta\gamma} = 2[\alpha\beta, \gamma\lambda] \quad (3.17)$$

subject at the same time to the symmetry requirements:

$$\begin{aligned} (a) \quad d_{\alpha\gamma, \beta\lambda} &= d_{\gamma\alpha, \beta\lambda}, \\ (b) \quad d_{\alpha\gamma, \beta\lambda} &= d_{\beta\lambda, \alpha\gamma}. \end{aligned} \quad (3.18)$$

We notice in the first place that if one only assumes the symmetry relations (a), the solution is already uniquely determined by (3.17). For, supposing that there are two such solutions: $d'_{\alpha\gamma, \beta\lambda}$ and $d''_{\alpha\gamma, \beta\lambda}$, their difference

$$\Delta d_{\alpha\gamma, \beta\lambda} = d'_{\alpha\gamma, \beta\lambda} - d''_{\alpha\gamma, \beta\lambda}$$

must clearly satisfy the following equation:

$$\Delta d_{\alpha\gamma, \beta\lambda} + \Delta d_{\alpha\lambda, \beta\gamma} = 0. \quad (3.19)$$

Re-labelling α and γ as γ and α , we can also write (3.19) as

$$\Delta d_{\gamma\alpha, \beta\lambda} + \Delta d_{\gamma\lambda, \beta\alpha} = 0. \quad (3.20)$$

Owing to the assumed symmetry between the first pair of indices, the first terms in (3.19) and (3.20) are equal. Hence, on subtracting (3.20) from (3.19), one obtains

$$\Delta d_{\alpha\lambda, \beta\gamma} - \Delta d_{\gamma\lambda, \beta\alpha} = 0. \quad (3.21)$$

Let us permute the first pair of indices in both terms, which is obviously permitted by the assumed symmetry, and afterwards re-label λ, α, γ as α, γ, λ , thus obtaining

$$\Delta d_{\alpha\gamma, \beta\lambda} - \Delta d_{\alpha\lambda, \beta\gamma} = 0. \quad (3.22)$$

Adding this equation to (3.19) we find

$$\Delta d_{\alpha\gamma, \beta\lambda} = 0. \quad (3.23)$$

$d'_{\alpha\gamma, \beta\lambda}$, $d''_{\alpha\gamma, \beta\lambda}$ are therefore necessarily equal, showing that with the assumed symmetry between the first pair of indices the solution of (3.17) is unique. It is easily verified that

$$d_{\alpha\gamma, \beta\lambda} = [\alpha\beta, \gamma\lambda] + [\beta\gamma, \alpha\lambda] - [\beta\lambda, \alpha\gamma] \quad (3.24)$$

satisfies (3.17) and, moreover, possesses the assumed symmetry between α and γ . Thus it is the only possible solution.

The solution is not in general compatible with the symmetry requirement (3.18b). In fact, on substituting (3.24) in (3.18b), a number of compatibility conditions on the square brackets are obtained. The interpretation is obvious. Since these results are obtained under the assumption that the reference configuration is free of stresses, these compatibility conditions clearly express the *necessary conditions* on the bracket expressions for vanishing stresses.

Let us now deduce these conditions explicitly. On substituting (3.24) in (3.18b), one obtains

$$[\beta\gamma, \alpha\lambda] - [\beta\lambda, \alpha\gamma] = [\alpha\lambda, \beta\gamma] - [\alpha\gamma, \beta\lambda]. \quad (3.25)$$

Re-labelling $\beta, \gamma, \alpha, \lambda$ as $\gamma, \beta, \lambda, \alpha$, one can write the relation also as

$$[\gamma\beta, \lambda\alpha] - [\gamma\alpha, \lambda\beta] = [\lambda\alpha, \gamma\beta] - [\lambda\beta, \gamma\alpha]. \quad (3.26)$$

Adding (3.25) and (3.26) and remembering the symmetry relations (3.12) for the square brackets, one obtains readily

$$[\beta\gamma, \alpha\lambda] = [\alpha\lambda, \beta\gamma]. \quad (3.27)$$

These are thus the necessary conditions for the reference configuration to be free of stresses.

The expressions for the elastic constants are obtained by substituting (3.24) in (3.16)*:

$$c_{\alpha\gamma, \beta\lambda} = [\alpha\beta, \gamma\lambda] + [\beta\gamma, \alpha\lambda] - [\beta\lambda, \alpha\gamma] + (\alpha\gamma, \beta\lambda). \quad (3.28)$$

It is, of course, understood that the bracket expressions are to be calculated for the reference configuration free of stresses. For this configuration (3.27) is satisfied; it is easily verified with the help of (3.27) that the $c_{\alpha\gamma, \beta\lambda}$ as given by (3.28) satisfy all the symmetry relations (3.1) as well as the equation (3.15).

4. COMPARISON WITH THE ELASTIC WAVE EQUATION IN A MEDIUM UNDER STRESS

The above method provides the general expressions for the elastic constants, but for a better understanding of the state of stresses and the significance of the conditions (3.27), a different procedure has to be followed. Thus we make no assumption

* The elastic constants are given in I and II as equal to $\{\alpha\gamma, \beta\lambda\}$ on the basis of the symmetry relations $\{\alpha\gamma, \beta\lambda\} = \{\gamma\alpha, \beta\lambda\} = \{\beta\lambda, \alpha\gamma\}$, which, as explained in §1, are not generally valid. $\{\alpha\gamma, \beta\lambda\}$ and $c_{\alpha\gamma, \beta\lambda}$ are equal only when the second and third terms accidentally cancel in (3.28), such as in the case of central interaction (§5) and probably also in other cases with particularly simple crystal symmetry.

about the reference configuration underlying the lattice wave equation (3.6) and compare the latter directly with the elastic wave equation in a medium which may be under any arbitrary system of homogeneous stresses.

In most standard treatises, general discussions of elasticity for media under stresses are not given.* The equations we shall require can, however, be obtained quite readily; the derivations are given in §6, the appendix. Let us consider a specimen which may initially be under any homogeneous stresses and denote the position vector of a point in it by \mathbf{x} . A homogeneous deformation is defined as

$$u_\alpha(\mathbf{x}) = \sum_\beta u_{\alpha\beta} x_\beta, \quad (4.1)$$

where $\mathbf{u}(\mathbf{x})$ is the displacement vector of a point at \mathbf{x} . The energy density due to the deformation is a function of $u_{\alpha\beta}$ and can be represented by a Taylor expansion

$$\Delta E = \sum_{\alpha\beta} S_{\alpha\beta} u_{\alpha\beta} + \frac{1}{2} \sum_{\alpha\gamma} \sum_{\beta\lambda} S_{\alpha\gamma, \beta\lambda} u_{\alpha\gamma} u_{\beta\lambda} + \dots, \quad (4.2)$$

where

$$S_{\alpha\gamma, \beta\lambda} = S_{\beta\lambda, \alpha\gamma}. \quad (4.3)$$

If the specimen in the deformed configuration is rigidly rotated, the energy density ΔE must obviously remain invariant. It is shown in the appendix that this invariance requires that

$$S_{\alpha\beta} = S_{\beta\alpha}, \quad (4.4)$$

$$S_{\alpha\lambda} \delta_{\beta\gamma} - S_{\gamma\lambda} \delta_{\beta\alpha} + S_{\alpha\gamma, \beta\lambda} - S_{\gamma\alpha, \beta\lambda} = 0. \quad (4.5)$$

For comparison with the lattice wave equation (3.6), we require only the elastic wave equation including up to quadratic terms of the wave number. The equation is obtained in the appendix and has a form very similar to the ordinary wave equation (3.5):

$$\rho \omega^2(y) u_\alpha = 4\pi^2 \sum_\beta \left(\sum_{\gamma\lambda} S_{\alpha\gamma, \beta\lambda} y_\gamma y_\lambda \right) u_\beta. \quad (4.6)$$

The difference from the above case is not so much in the wave equation as in the different symmetry relations (4.3), (4.4), (4.5) for $S_{\alpha\gamma, \beta\lambda}$. $S_{\alpha\beta} = S_{\beta\alpha}$, we notice, are the existing stresses in the 'undeformed' specimen. If we equate them to zero, $S_{\alpha\gamma, \beta\lambda}$ should become the ordinary elastic constants $c_{\alpha\gamma, \beta\lambda}$. It is seen that (4.3), (4.5) actually reduce, in that case, to the symmetry relations (3.1) for $c_{\alpha\gamma, \beta\lambda}$.

The comparison of (4.6) with the lattice wave equation (3.6) leads to

$$\begin{aligned} S_{\alpha\gamma, \beta\lambda} + S_{\alpha\lambda, \beta\gamma} &= \{\alpha\gamma, \beta\lambda\} + \{\alpha\lambda, \beta\gamma\} \\ &= 2[\alpha\beta, \gamma\lambda] + (\alpha\gamma, \beta\lambda) + (\alpha\lambda, \beta\gamma). \end{aligned} \quad (4.7)$$

(4.7), together with the symmetry relations (4.3), (4.4), (4.5), can be regarded as a system of equations determining $S_{\alpha\beta}$, $S_{\alpha\gamma, \beta\lambda}$. It is obvious from the special example given in §2 that the solution cannot be completely determinate. For in that case, all the stresses $S_{\alpha\beta}$ would be explicitly expressed, which directly contradicts the special example. In fact, we shall find that apart from $S_{11} + S_{22} + S_{33}$ which is perfectly arbitrary, the values of $S_{\alpha\beta}$ are otherwise completely determined by the equations.

* In fact, the only discussion of this type I have found (Brillouin 1946, p. 232) is apparently in error precisely in parts which would concern the following considerations.

It is convenient to introduce instead of $S_{\alpha\gamma, \beta\lambda}$, $x_{\alpha\beta, \gamma\lambda}$ as unknowns which are defined by

$$S_{\alpha\gamma, \beta\lambda} = \{[\alpha\beta, \gamma\lambda] + [\beta\gamma, \alpha\lambda] - [\beta\lambda, \alpha\gamma] + (\alpha\gamma, \beta\lambda)\} + x_{\alpha\beta, \gamma\lambda}. \quad (4.8)$$

It has already been shown in the preceding sections that the four terms enclosed in the curly brackets satisfy equations of the form (4.7), and, moreover, taken together, are symmetric in $\alpha\gamma$. Hence on expressing $S_{\alpha\gamma, \beta\lambda}$ in terms of $x_{\alpha\beta, \gamma\lambda}$, we find that (4.7) reduces to the requirement that $x_{\alpha\beta, \gamma\lambda}$ is antisymmetric in the last pair of indices

$$x_{\alpha\beta, \gamma\lambda} = -x_{\alpha\beta, \lambda\gamma}. \quad (4.9)$$

In terms of $x_{\alpha\beta, \gamma\lambda}$ the symmetry relations (4.5) remain relatively simple:

$$S_{\alpha\lambda} \delta_{\beta\gamma} - S_{\gamma\lambda} \delta_{\beta\alpha} + x_{\alpha\beta, \gamma\lambda} - x_{\gamma\beta, \alpha\lambda} = 0. \quad (4.10)$$

The simple relations (4.3), on the other hand, assume the more complicated appearance

$$x_{\alpha\beta, \gamma\lambda} + [\beta\gamma, \alpha\lambda] - [\beta\lambda, \alpha\gamma] = x_{\beta\alpha, \lambda\gamma} + [\alpha\lambda, \beta\gamma] - [\alpha\gamma, \beta\lambda]. \quad (4.11)$$

If in (4.10) we put

$$\beta = \gamma \neq \alpha,$$

the relation reduces to

$$S_{\alpha\lambda} = x_{\beta\beta, \alpha\lambda} - x_{\alpha\beta, \beta\lambda} \quad (\alpha \neq \beta). \quad (4.12)$$

The expression on the right must obviously be independent of β so long as $\beta \neq \alpha$. For $\lambda \neq \alpha$ we can put $\beta = \lambda$, hence

$$S_{\alpha\lambda} = x_{\lambda\lambda, \alpha\lambda} - x_{\alpha\lambda, \lambda\lambda} = x_{\lambda\lambda, \alpha\lambda} \quad (\alpha \neq \lambda), \quad (4.13)$$

where $x_{\alpha\lambda, \lambda\lambda}$ vanishes owing to the antisymmetry requirement of (4.9). For $\alpha = \lambda$, the first term in (4.12) vanishes for the same reason, leaving

$$S_{\alpha\alpha} = -x_{\alpha\beta, \beta\alpha} \quad (\alpha \neq \beta). \quad (4.14)$$

This implies of course that $x_{\alpha\beta, \beta\alpha}$ is independent of β , if $\beta \neq \alpha$.

Now that $S_{\alpha\beta}$ are expressed in terms of $x_{\alpha\beta, \gamma\lambda}$, let us proceed to obtain as far as possible the solutions for the latter. Clearly, at least two of the indices in $x_{\alpha\beta, \gamma\lambda}$ will be identical; we shall obtain the solutions separately for different positions of the identical indices:

(a) $x_{\alpha\beta, \gamma\gamma}$: It follows directly from the antisymmetry requirement (4.9) that

$$x_{\alpha\beta, \gamma\gamma} = 0. \quad (4.15)$$

(b) $x_{\alpha\alpha, \gamma\lambda}$: On putting $\beta = \alpha$ in (4.11) we obtain, with the help of (4.9), that

$$x_{\alpha\alpha, \gamma\lambda} = [\alpha\lambda, \alpha\gamma] - [\alpha\gamma, \alpha\lambda]. \quad (4.16)$$

(c) $x_{\alpha\beta, \beta\lambda}$: If we restrict ourselves to the case $\alpha \neq \lambda$, $\alpha \neq \beta$, we can combine (4.12) with (4.13) and obtain

$$\begin{aligned} x_{\alpha\beta, \beta\lambda} &= x_{\beta\beta, \alpha\lambda} - S_{\alpha\lambda} \\ &= x_{\beta\beta, \alpha\lambda} - x_{\lambda\lambda, \alpha\lambda} \quad (\alpha \neq \lambda, \alpha \neq \beta). \end{aligned}$$

Both terms are of the type given under (b) and may hence be written in terms of the square brackets by suitable re-labelling of the indices in (4.16):

$$x_{\alpha\beta, \beta\lambda} = [\beta\lambda, \beta\alpha] - [\beta\alpha, \beta\lambda] - [\lambda\lambda, \lambda\alpha] + [\lambda\alpha, \lambda\lambda] \quad (\alpha \neq \lambda, \alpha \neq \beta). \quad (4.17)$$

(d) $x_{\alpha\beta, \gamma\beta}$: Owing to the antisymmetry in the last pair of indices,

$$x_{\alpha\beta, \gamma\beta} = -x_{\alpha\beta, \beta\gamma}.$$

If $\alpha \neq \beta$, $\alpha \neq \gamma$, this case is immediately reduced to case (c) by re-labelling λ as γ in (4.17).

$$x_{\alpha\beta, \gamma\beta} = -x_{\alpha\beta, \beta\gamma} = -[\beta\gamma, \beta\alpha] + [\beta\alpha, \beta\gamma] + [\gamma\gamma, \gamma\alpha] - [\gamma\alpha, \gamma\gamma] \quad (\alpha \neq \beta, \alpha \neq \gamma). \quad (4.18)$$

(e) $x_{\alpha\beta, \gamma\alpha}$: For this case, we obtain, on putting $\lambda = \alpha$ in (4.11),

$$x_{\alpha\beta, \gamma\alpha} + [\beta\gamma, \alpha\alpha] - [\beta\alpha, \alpha\gamma] = x_{\beta\alpha, \alpha\gamma} + [\alpha\alpha, \beta\gamma] - [\alpha\gamma, \beta\alpha].$$

If we consider only the case $\beta \neq \alpha$, $\beta \neq \gamma$, $x_{\beta\alpha, \alpha\gamma}$ becomes of the type given in (c) and may be eliminated after re-labelling α, β, λ in (4.12) as β, α, γ . Thus we obtain

$$x_{\alpha\beta, \gamma\alpha} = [\alpha\alpha, \beta\gamma] - [\beta\gamma, \alpha\alpha] - [\gamma\gamma, \gamma\beta] + [\gamma\beta, \gamma\gamma] \quad (\beta \neq \alpha, \beta \neq \gamma). \quad (4.19)$$

(f) $x_{\alpha\beta, \alpha\lambda}$: For $\beta \neq \alpha$, $\beta \neq \lambda$, this case is reduced to that above by the antisymmetry relation

$$x_{\alpha\beta, \alpha\lambda} = -x_{\alpha\beta, \lambda\alpha} = -[\alpha\alpha, \beta\lambda] + [\beta\lambda, \alpha\alpha] + [\lambda\lambda, \lambda\beta] - [\lambda\beta, \lambda\lambda] \quad (\beta \neq \alpha, \beta \neq \lambda), \quad (4.20)$$

where (4.19) is used with γ re-labelled as λ .

The cases listed above (the inequality restrictions on the indices in (4.17) to (4.20) are particularly to be noticed) cover, in fact, consistently and completely all possible forms of $x_{\alpha\beta, \gamma\lambda}$ except the following:

$$x_{\alpha\beta, \beta\alpha} = -x_{\alpha\beta, \alpha\beta} \quad (\alpha \neq \beta).$$

This case is related to the stresses (4.14) and cannot be completely expressed in terms of the square brackets. The various individual cases are related by

$$\left. \begin{aligned} S_{11} &= -x_{12, 21} = -x_{13, 31} = x_{12, 12} = x_{13, 13}, \\ S_{22} &= -x_{23, 32} = -x_{21, 12} = x_{23, 23} = x_{21, 21}, \\ S_{33} &= -x_{31, 13} = -x_{32, 23} = x_{31, 31} = x_{32, 32}. \end{aligned} \right\} \quad (4.21)$$

The differences between these stresses are determined by (4.11). For we can write

$$S_{11} - S_{22} = x_{21, 12} - x_{12, 21}, \quad (4.22)$$

$$S_{22} - S_{33} = x_{32, 23} - x_{23, 32}, \quad (4.23)$$

$$S_{33} - S_{11} = x_{13, 31} - x_{31, 13}. \quad (4.24)$$

The right-hand sides of the equations may be expressed in terms of the square brackets by putting $\alpha = \lambda = 2$, $\beta = \gamma = 1$; $\alpha = \lambda = 3$, $\beta = \gamma = 2$; and $\alpha = \lambda = 1$, $\beta = \gamma = 3$, respectively in (4.11). One obtains in this way:

$$S_{11} - S_{22} = [22, 11] - [11, 22], \quad (4.25)$$

$$S_{22} - S_{33} = [33, 22] - [22, 33], \quad (4.26)$$

$$S_{33} - S_{11} = [11, 33] - [33, 11]. \quad (4.27)$$

Let us introduce the average pressure

$$p = -\frac{1}{3}(S_{11} + S_{22} + S_{33}), \quad (4.28)$$

which we find not expressible in the general theory. Combining (4.28) with (4.25), (4.27), one finds that

$$S_{11} = -p + \frac{1}{3}([22, 11] + [33, 11] - [11, 22] - [11, 33]). \quad (4.29)$$

S_{22} , S_{33} can be obtained by cyclic permutations. With the help of (4.21) all the quantities of the form $x_{\alpha\beta, \beta\alpha} = -x_{\alpha\beta, \alpha\beta}$ can be similarly expressed in terms of p and the square brackets.

We shall complete the description of the solutions by giving $S_{\alpha\lambda} (\alpha \neq \lambda)$ in terms of the square brackets. Let us put $\alpha = \lambda$ in (4.16), and then re-label γ as α . By (4.13) the result is equal to $S_{\alpha\lambda}$, hence

$$S_{\alpha\lambda} = [\lambda\lambda, \lambda\alpha] - [\lambda\alpha, \lambda\lambda] \quad (\alpha \neq \lambda). \quad (4.30)$$

Since $S_{\alpha\gamma, \beta\lambda}$ is expressible in terms of $x_{\alpha\beta, \gamma\lambda}$ by (4.8), we have thus expressed all $S_{\alpha\beta}$, $S_{\alpha\beta, \beta\lambda}$ in terms of p and the brackets. These will be the solutions of (4.4), (4.5), (4.7) only under the assumption that a solution exists. This has to be verified by substituting the above expression for $S_{\alpha\beta}$, $S_{\alpha\gamma, \beta\lambda}$ in the equations. In fact, the equations are not identically satisfied in this way, one obtains instead ten identities which are thus the conditions for the existence of the solution. Since it is obvious, from physical grounds, that the solutions must exist, these identities must necessarily be satisfied by the brackets. The identities are

$$[22, 11] + [33, 22] + [11, 33] = [11, 22] + [22, 33] + [33, 11], \quad (4.31)$$

$$[21, 13] = [31, 12] \quad \text{and two more by cyclic permutations,} \quad (4.32)$$

$$[22, 23] - [23, 22] = [33, 32] - [32, 33] \quad \text{and cyclic permutations,} \quad (4.33)$$

$$[33, 23] - [23, 33] = [11, 23] - [23, 11] \quad \text{and cyclic permutations.} \quad (4.34)$$

Since they have been obtained without making any special assumptions apart from the rotational invariance condition in the elastic theory expressed by (4.4) and (4.5), they must clearly express the corresponding invariance conditions in the lattice theory. These relations are thus automatically fulfilled in any special example.

The interpretation of the fifteen necessary conditions (3.27) for vanishing stresses is now obvious. We notice that if they are fulfilled, it follows from (4.30) that

$$S_{\alpha\lambda} = 0 \quad (\alpha \neq \lambda)$$

and from (4.25) and (4.27) that

$$S_{11} - S_{22} = S_{33} - S_{11} = 0.$$

These are thus sufficient conditions for all the anisotropic stresses to vanish, leaving undetermined only an isotropic pressure:

$$S_{11} = S_{22} = S_{33} = -p.$$

Moreover, the ten invariance relations (4.31) to (4.34) are seen to be identically satisfied in this case. Thus what have been described as the necessary conditions for vanishing stresses are in fact equivalent to the five conditions for vanishing anisotropic stresses plus the ten invariance conditions. Since the latter are automatically fulfilled in any case, only five relations are actually imposed by (3.27), which clearly

determine the shape of the lattice cell. Only in a special example can the condition for vanishing pressure be formulated, which will determine the equilibrium size of the lattice cell.

5. CENTRAL FORCES

To illustrate the above results, it is sufficient to apply the theory to cases where each lattice point is a centre of symmetry. For the latter assumption only makes the round brackets (3.9) vanish (see below), whereas the various identities and conditions obtained in the last two sections concern only the square brackets (3.8).

One consequence of the assumption of periodicity is that the second derivative

$$\left(\frac{\partial^2 \Phi}{\partial u_\alpha \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) \partial u_\beta \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)} \right)_0$$

depends on the lattice cell indices l, l' only through the difference $l - l'$. Thus in the general theory it is denoted by

$$\Phi_{\alpha\beta} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) = \left(\frac{\partial^2 \Phi}{\partial u_\alpha \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) \partial u_\beta \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)} \right)_0. \quad (5.1)$$

In (3.8) the derivatives appear in this form; the remaining quantities v_α and $\mathbf{x} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)$ are respectively the lattice cell volume and the relative lattice vector defined by

$$\mathbf{x} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) = \mathbf{x} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - \mathbf{x} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right). \quad (5.2)$$

Let $\phi_{kk'}(r)$ be the central potential between lattice particles of the types k and k' . The second derivatives can be obtained by differentiating the formal expression of the total potential Φ , namely,

$$\Phi = \frac{1}{2} \sum_{kk'} \sum_{l'k'}' \phi_{kk'} \left(\left| \mathbf{x} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) + \mathbf{u} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - \mathbf{u} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right) \right| \right), \quad (5.3)$$

where the prime excludes $\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right) = \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)$, and the absolute magnitude of the vector indicated is always taken for the argument which is, of course, just the distance between $\left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)$ and $\left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)$ in their displaced positions. Remembering that

$$\begin{aligned} \frac{\partial}{\partial u_\alpha \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right)} \left| \mathbf{x} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) + \mathbf{u} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - \mathbf{u} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right) \right| &= - \frac{\partial}{\partial u_\alpha \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)} \left| \mathbf{x} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) + \mathbf{u} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - \mathbf{u} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right) \right| \\ &= \frac{x_\alpha \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) + u_\alpha \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - u_\alpha \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right)}{\left| \mathbf{x} \left(\begin{smallmatrix} l-l' \\ k \quad k' \end{smallmatrix} \right) + \mathbf{u} \left(\begin{smallmatrix} l \\ k \end{smallmatrix} \right) - \mathbf{u} \left(\begin{smallmatrix} l' \\ k' \end{smallmatrix} \right) \right|}, \end{aligned}$$

one finds readily that for $\begin{pmatrix} l \\ k \end{pmatrix} \neq \begin{pmatrix} 0 \\ k' \end{pmatrix}$

$$\Phi_{\alpha\beta}\left(\begin{matrix} l \\ kk' \end{matrix}\right) = -\delta_{\alpha\beta} D\phi_{kk'}\left(\left|\mathbf{x}\left(\begin{matrix} l \\ kk' \end{matrix}\right)\right|\right) - x_{\alpha}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\beta}\left(\begin{matrix} l \\ kk' \end{matrix}\right) D^2\phi_{kk'}\left(\left|\mathbf{x}\left(\begin{matrix} l \\ kk' \end{matrix}\right)\right|\right), \quad (5.4)$$

where D is the operator (for this and other details, cf. series of papers headed by Born 1940)

$$D = \frac{1}{r} \frac{d}{dr}. \quad (5.5)$$

Before using (5.4) to discuss the general result, one notices that the round brackets vanish in this case because it follows from the definition that

$$\Phi_{\alpha\beta}\left(\begin{matrix} l \\ kk' \end{matrix}\right) = \Phi_{\beta\alpha}\left(\begin{matrix} -l \\ k'k \end{matrix}\right). \quad (5.6)$$

For then if every particle is a centre of symmetry, any other particle in the lattice will have an identical particle diagonally on the other side; hence, in a summation of the type

$$\sum_{lk'} \Phi_{\alpha\mu}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\gamma}\left(\begin{matrix} l \\ kk' \end{matrix}\right),$$

their contributions cancel by pairs. Physically this result is obvious. The round bracket represents the effect of the shift of the particles in the basis during a homogeneous deformation (cf. II). If every lattice particle is a centre of symmetry, obviously no such shifts will occur.

Substituting (5.4) in (3.8), one obtains

$$\begin{aligned} [\alpha\beta, \gamma\lambda] &= \frac{1}{2v_a} \delta_{\alpha\beta} \sum_{lkk'} x_{\gamma}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\lambda}\left(\begin{matrix} l \\ kk' \end{matrix}\right) D\phi_{kk'}\left(\left|\mathbf{x}\left(\begin{matrix} l \\ kk' \end{matrix}\right)\right|\right) \\ &\quad + \frac{1}{2v_a} \sum_{lkk'} x_{\alpha}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\beta}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\gamma}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\lambda}\left(\begin{matrix} l \\ kk' \end{matrix}\right) D^2\phi_{kk'}\left(\left|\mathbf{x}\left(\begin{matrix} l \\ kk' \end{matrix}\right)\right|\right). \end{aligned} \quad (5.7)$$

The ten identities (4.31) to (4.34) are immediately seen to be satisfied. The second term in (5.7) is symmetric in all four indices and consequently drops out of these identities directly. In (4.31) the δ -symbol becomes unity in all terms and the two sides become the identical sum of three terms. (4.32), owing to the δ -symbol, reduces simply to $0 = 0$. In (4.33), only the first terms on the two sides survive; in these terms the δ -symbols become unity and the remainders are obviously identical. Exactly the same argument applies to (4.34).

The shear stresses in the co-ordinate directions $S_{\alpha\lambda} (\alpha \neq \lambda)$ are given by substituting (5.7) in (4.30):

$$S_{\alpha\lambda} = \frac{1}{2v_a} \sum_{lkk'} x_{\alpha}\left(\begin{matrix} l \\ kk' \end{matrix}\right) x_{\lambda}\left(\begin{matrix} l \\ kk' \end{matrix}\right) D\phi_{kk'}\left(\left|\mathbf{x}\left(\begin{matrix} l \\ kk' \end{matrix}\right)\right|\right), \quad (5.8)$$

which agree with the expressions for the stresses given by Love (1944) for a simple lattice, and by Born (1923) for the general case.

To give the tensions in the co-ordinate directions, one has to work out the average pressure from the given potential. For this purpose, consider an isotropic expansion

$$\mathbf{u}\left(\begin{matrix} l \\ k \end{matrix}\right) = \mathbf{x}\left(\begin{matrix} l \\ k \end{matrix}\right) \delta. \quad (5.9)$$

The corresponding energy per unit cell can be written as

$$\Delta E = \frac{1}{2} \sum_{k'} \sum_{\substack{(l) \\ (k)}} \phi_{kk'} \left((1 + \delta) \left| \mathbf{x} \left(\frac{l}{kk'} \right) \right| \right). \quad (5.10)$$

During this expansion each edge of a unit cube is increased by δ in length; hence macroscopically the work done on the cube is

$$(S_{11} + S_{22} + S_{33}) \delta = -3p\delta. \quad (5.11)$$

This is equal to the energy increase in unit volume which can be obtained by dividing ΔE by v_a . Hence

$$-3p = \frac{1}{v_a} \left(\frac{d\Delta E}{d\delta} \right)_0 = \frac{1}{2v_a} \sum_{kk'} \left| \mathbf{x} \left(\frac{l}{kk'} \right) \right|^2 D\phi_{kk'} \left(\left| \mathbf{x} \left(\frac{l}{kk'} \right) \right| \right). \quad (5.12)$$

Using this value for p in (4.29), one finds

$$\begin{aligned} S_{11} &= -p + \frac{1}{6v_a} \sum_{lkk'} \left(2x_1^2 \left(\frac{l}{kk'} \right) - x_2^2 \left(\frac{l}{kk'} \right) - x_3^2 \left(\frac{l}{kk'} \right) \right) D\phi_{kk'} \left(\left| \mathbf{x} \left(\frac{l}{kk'} \right) \right| \right) \\ &= \frac{1}{2v_a} \sum_{lkk'} x_1^2 \left(\frac{l}{kk'} \right) D\phi_{kk'} \left(\left| \mathbf{x} \left(\frac{l}{kk'} \right) \right| \right), \end{aligned} \quad (5.13)$$

which again agrees with the results of Born and Love.

To obtain the elastic constants, it must be required that all the stresses vanish. It follows from (5.13) and (5.8) that the first term in $[\alpha\beta, \gamma\lambda]$ vanishes. $[\alpha\beta, \gamma\lambda]$ thus becomes symmetric in all four indices. Furthermore, as the round brackets vanish, the expression (3.28) for the elastic constants reduces to

$$c_{\alpha\gamma, \beta\lambda} = [\alpha\beta, \gamma\lambda]. \quad (5.14)$$

The six relations expressing the complete symmetry in the four tensor indices are the Cauchy relations, usually expressed in Voigt's notation as

$$\begin{aligned} c_{23} &= c_{44}, & c_{31} &= c_{55}, & c_{12} &= c_{66}, \\ c_{14} &= c_{56}, & c_{25} &= c_{46}, & c_{36} &= c_{45}. \end{aligned}$$

Recently, Epstein (1946) has expressed doubts concerning the derivation of these relations. Zener (1947) has later emphasized that these relations are valid under three assumptions: (i) central forces, (ii) every point a centre of symmetry, (iii) vanishing initial stresses.* The way in which (ii) and (iii) affect the Cauchy relations is clear from the above derivations; the presence either of the round bracket (ii) or the first term in (5.7) would invalidate these relations. It appears, however, that (iii) should be considered a prerequisite for any definition of the elastic constants. The symmetry relations (4.5) in the more general case with stresses show that $S_{\alpha\gamma, \beta\lambda}$ do not satisfy the usual symmetry relations (3.1) for the elastic constants.

* This condition is apparently not heeded in tables II, III given on p. 236 of Brillouin's book (Brillouin 1946), where the number of 'elastic constants' has been reduced by 6 with Cauchy's relations for crystals under stresses.

Voigt's notation would, for instance, become quite inapplicable. Thus one could hardly regard $S_{\alpha\gamma, \beta\gamma}$ as the elastic constants in the usual sense and (iii) cannot be considered a special restriction applying only to the Cauchy relations.

6. APPENDIX ELASTICITY FOR MEDIA UNDER ARBITRARY HOMOGENEOUS STRESSES

The coefficients in the expressions (4.2) for the energy of deformation are not completely independent owing to the requirement of rotational invariance. Let us subject the arbitrarily deformed system ($u_{\alpha\beta}$) to an infinitesimal rotation rigidly. A point originally at \mathbf{x} has the position vector

$$x_\alpha + \sum_\beta u_{\alpha\beta} x_\beta \quad (6.1)$$

in the deformed configuration. If the infinitesimal rotation is specified by the antisymmetric matrix

$$\omega_{\alpha\beta} = -\omega_{\beta\alpha}, \quad (6.2)$$

the same point after the rotation is moved into

$$x_\alpha + \sum_\beta u_{\alpha\beta} x_\beta + \sum_\beta \omega_{\alpha\beta} (x_\beta + \sum_\gamma u_{\beta\gamma} x_\gamma). \quad (6.3)$$

Since this is still linear in the original co-ordinates x_α , the combined deformation and rotation is once more a homogeneous deformation of the type (4.1), in fact, with the deformation components

$$\overline{u}_{\alpha\beta} = u_{\alpha\beta} + \omega_{\alpha\beta} + \sum_\gamma \omega_{\alpha\gamma} u_{\gamma\beta}. \quad (6.4)$$

The energy density corresponding to this state is obtained by substituting this expression in the expansion (4.2), thus

$$\Delta E = \sum_{\alpha\beta} (u_{\alpha\beta} + \omega_{\alpha\beta}) S_{\alpha\beta} + \sum_{\alpha\beta\gamma} \omega_{\alpha\gamma} u_{\gamma\beta} S_{\alpha\beta} + \frac{1}{2} \sum_{\alpha\beta} \sum_{\gamma\lambda} (u_{\alpha\beta} + \omega_{\alpha\beta}) (u_{\gamma\lambda} + \omega_{\gamma\lambda}) S_{\alpha\beta, \gamma\lambda} + \dots \quad (6.5)$$

The rotational invariance requires that this expression is to the first order independent of $\omega_{\alpha\beta}$. Hence (6.5) may be differentiated with respect to $\omega_{\alpha\beta}$ ($\omega_{\alpha\beta} = -\omega_{\beta\alpha}$!) and the result set equal to zero:

$$S_{\alpha\beta} - S_{\beta\alpha} + \sum_\mu \{u_{\beta\mu} S_{\alpha\mu} - u_{\alpha\mu} S_{\beta\mu}\} + \sum_{\gamma\lambda} (S_{\alpha\beta, \gamma\lambda} - S_{\beta\alpha, \gamma\lambda}) u_{\gamma\lambda} + \text{higher terms in } u_{\alpha\beta} = 0. \quad (6.6)$$

This relation obviously holds for any arbitrary values of $u_{\alpha\beta}$; hence the various coefficients must identically vanish:

$$S_{\alpha\beta} = S_{\beta\alpha}, \quad (6.7)$$

$$S_{\alpha\lambda} \delta_{\beta\gamma} - S_{\beta\lambda} \delta_{\alpha\gamma} + S_{\alpha\beta, \gamma\lambda} - S_{\beta\alpha, \gamma\lambda} = 0. \quad (6.8)$$

(6.8) is most easily obtained in this form by differentiation with respect to $u_{\gamma\lambda}$. (4.5) used in the text is the same as (6.8) with β, γ re-labelled as γ, β .

On p. 232 of Brillouin's book (Brillouin 1946) $S_{\alpha\gamma, \beta\lambda}$ are stated to fulfil the symmetry relations (3.1) of the ordinary elastic constants, for any state of initial stresses

$S_{\alpha\beta}$. In view of (6.8), this is apparently an error. In fact, had $S_{\alpha\gamma, \beta\lambda}$ satisfied the relations (3.1), the comparison in §4 would become identical with §3 with the consequent requirement of the fifteen conditions (3.27), which would, as we have seen, give vanishing values for all the anisotropic stresses.

With the given energy of deformation the wave equation we require can be obtained by constructing the Lagrangian

$$L = \int \left\{ \frac{1}{2} \rho |\dot{\mathbf{u}}(\mathbf{x})|^2 - \Delta E \left(\frac{\partial u_\beta(\mathbf{x})}{\partial x_\alpha} \right) \right\} d\mathbf{x}. \quad (6.9)$$

From the Lagrangian the Euler equations can be obtained by applying the variational principle in the usual way:

$$\begin{aligned} \rho \ddot{u}_\alpha(\mathbf{x}) &= \sum_\gamma \frac{\partial}{\partial x_\gamma} \left\{ S_{\alpha\gamma} + \frac{1}{2} \sum_{\beta\lambda} S_{\alpha\gamma, \beta\lambda} \frac{\partial u_\beta}{\partial x_\lambda} + \frac{1}{2} \sum_{\beta\lambda} S_{\beta\lambda, \alpha\gamma} \frac{\partial u_\beta}{\partial x_\lambda} \right\} \\ &= \sum_\beta \left(\sum_{\gamma\lambda} S_{\alpha\gamma, \beta\lambda} \frac{\partial u_\beta(\mathbf{x})}{\partial x_\gamma \partial x_\lambda} \right). \end{aligned} \quad (6.10)$$

For a plane wave

$$\mathbf{u} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x} - i\omega t\}, \quad (6.11)$$

(6.10) reduces to

$$\rho \omega^2 u_\alpha = 4\pi^2 \sum_\beta \left(\sum_{\gamma\lambda} S_{\alpha\gamma, \beta\lambda} y_\gamma y_\lambda \right) u_\beta. \quad (6.12)$$

In the wave, the deformation is, of course, nowhere uniform, and the energy density will in general be a function of the higher derivatives of the elastic displacement \mathbf{u} . We could, for instance, represent the energy density by an expansion in terms of all the derivatives. However, by working out one or two of the lowest terms, such as the mixed term $(\partial u_\beta / \partial x_\gamma)_0 (\partial^2 u_\gamma / \partial x_\lambda \partial x_\mu)_0$, one immediately finds that they will only give rise to terms in the wave equation higher than the quadratic in the wave-number components. Since the lattice wave equation (3.6) only includes up to the quadratic terms, the effects of the non-homogeneous strains are irrelevant for the present discussion.

REFERENCES

- Begbie, G. H. & Born, M. 1947 *Proc. Roy. Soc. A*, **188**, 179.
 Born, M. 1923 *Atomtheorie des festen Zustandes*. Berlin: Teubner.
 Born, M. 1940 *Proc. Camb. Phil. Soc.* **36**, 160.
 Born, M. 1942 *Rep. Progr. Phys.* **9**, 294.
 Born, M. 1945 *Rev. Mod. Phys.* **17**, 245.
 Born, M. & Goppert-Mayer, M. 1933 *Handb. Phys.* **24** (2nd part), 623. Berlin: Springer.
 Brillouin, L. 1946 *Les tenseurs en mécanique et en élasticité*. New York: Dover.
 Epstein, P. S. 1946 *Phys. Rev.* **70**, 915.
 Huang, K. 1949 *Phil. Mag.* **40**, 733.
 Love, A. E. H. 1944 *A treatise on the mathematical theory of elasticity*, p. 619. New York: Dover.
 Seitz, F. 1940 *Modern theory of solids*, chap. 14. New York: McGraw Hill.
 Smith, Helen M. J. 1948 *Phil. Trans. A*, **241**, 105.
 Voigt, W. 1928 *Lehrbuch der Kristallphysik*. Berlin: Teubner.
 Zener, C. 1947 *Phys. Rev.* **71**, 323.

The surface impedance of superconductors and normal metals at high frequencies

V. Analysis of experimental results for superconducting tin

By A. B. PIPPARD, *The Royal Society Mond Laboratory, University of Cambridge*

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 16 March 1950)

For the purpose of analyzing the experimental results presented in the earlier parts of this series the concept of a two-parameter two-fluid model of superconductivity is developed, and it is shown that current theories are special cases of this model. The greater part of the analysis is carried out by dimensional arguments based on the model. The part played by the resistive mechanism in modifying the surface reactance is discussed, and it is concluded that the temperature variation of inductive skin depth at the lowest temperatures represents the variation of the superconducting penetration depth. The dependence of this quantity on crystal orientation is different from what would be expected from the theory of London & London, and it is concluded that there may be grounds for doubting the validity of this theory, or at least its application to high-frequency phenomena. The dimensional analysis of the resistance measurements shows them to fit reasonably well into the framework of the two-fluid model, but the implications of the particular form of the model which is deduced from the results are in disagreement with other general considerations. In particular, the resistance is found to vary too slowly with frequency and with the fraction of normal electrons. The conclusion is drawn that the two-fluid model as formulated here does not give a satisfactory account of the phenomena, and that the detailed theories which belong to this category therefore also fail.

THE TWO-FLUID MODEL OF SUPERCONDUCTIVITY

When we attempt to analyze the results of measurements of the high-frequency impedance of superconducting tin described in the previous papers of this series (Pippard 1947, 1950), we must start with some model of a superconductor. Now there have been many theoretical treatments of superconductivity, ranging from the purely phenomenological theory of London & London (1935) to the more detailed electron theory of Heisenberg (1947), and it is not convenient to work out completely the high-frequency effects to be expected in each model. We shall therefore begin by looking for some common principles underlying all these theories which may be used as a starting-point for the analysis. Naturally, the more theories we embrace in our treatment the less detailed must the treatment be, and in fact what we shall attempt here is to find a basis for a dimensional analysis of the experimental results. The extent to which the results agree with the dimensional analysis may be taken as a measure of the validity of the basic principles of existing theories of superconductivity.

So far as high-frequency phenomena are concerned the theory of London & London yields the same results as the earlier acceleration theory of Becker, Heller & Sauter (1933), which was extended by H. London (1934) to include resistive effects and used by him in discussing his measurements on tin at 1500 Mc./sec. (H. London 1940). He assumed that the electrons in a superconductor might be thought of as belonging to either of two independent groups, superconducting and normal respectively. The

current density \mathbf{J} at any point could thus be considered as the sum of contributions, \mathbf{J}_s and \mathbf{J}_n , from each group, and the equations governing the dependence of each current density on the electric field \mathbf{E} were written as follows:

$$\Lambda \mathbf{J}_s = \mathbf{E}, \quad (1)$$

and
$$\mathbf{J}_n = \sigma \mathbf{E}, \quad (2)$$

where σ is the 'effective conductivity' of the normal electrons, not necessarily the same as the conductivity of the normal metal. With these and Maxwell's equations it is a straightforward matter to calculate the surface impedance at any frequency as a function of the two parameters Λ and σ .

Now we saw in II that the occurrence of the anomalous skin effect at low temperatures renders equation (2) incorrect, and in fact the dependence of \mathbf{J}_n on \mathbf{E} is governed not by a single parameter, σ , but also by l , the mean free path of the conduction electrons. When, however, the conditions in a normal metal are such as to make l very much greater than the depth of penetration of the field into the metal, the resistance becomes independent of σ , and the current is again governed by only one parameter, σ/l , which is given for a free-electron model by ne^2/mv . This result was applied semi-quantitatively in III to a superconductor by the use of the ineffectiveness concept, and since then the detailed theory has been worked out by Sondheimer (unpublished)* and by Maxwell, Marcus & Slater (1949), using equation (1) for the supercurrent and the appropriate integro-differential equation (Reuter & Sondheimer 1948, equation (14)) for the normal current. These calculations show that in the superconductor as well as in the normal metal it is the ratio σ/l which determines the behaviour of \mathbf{J}_n , provided that l is much greater than the skin depth, thus the surface impedance is again governed by only two parameters, Λ and σ/l .

These models are typical of what we shall refer to as 'two-fluid models' of a superconductor, by analogy with the similar model of liquid helium. From the point of view of the analysis which will follow the essential feature of the models is that already remarked on, that the electrodynamic equation may be broken up into two separate equations, describing the supercurrent and the normal current independently. Apart from this the equations, if they are to account for the experimental results, must be linear, since the impedance has been found to be independent of the strength of the measuring current, for the small currents used in this work at any rate. The particular two-fluid models which we have considered so far also have another property, that only one parameter is needed to describe each contribution to the total current, and these models may therefore be termed 'two-parameter two-fluid models'.

There is some justification for the view that superconductivity may be described by such a two-fluid model, in that it was shown by Gorter & Casimir (1934) that the thermodynamic properties of a superconductor may be adequately described by writing the free energy of the electron assembly as the sum of two contributions, due to the superconducting and normal electrons respectively. Moreover, the variation with temperature of the number of super-electrons in this representation agrees well,

* I am grateful to Dr E. H. Sondheimer for making available these unpublished calculations.

as was pointed out by Daunt, Miller, Pippard & Shoenberg (1948), with the observed variation of the penetration depth. Thus the thermodynamic and magnetostatic properties of superconductors may be consistently described by a two-fluid model which combines the equations of Gorter & Casimir and of London & London. The importance of measurements at high frequencies lies in the fact that they enable this scheme to be subjected to further experimental test, since, alone among experiments on superconductors, they give fairly direct evidence concerning the behaviour of the normal electrons.

Now the models discussed so far are very little further advanced than phenomenological theories, no attempt having been made by their authors to show how electronic interactions in real metals might give rise to the characteristic two-fluid behaviour. Recently, however, there have been proposed two much more detailed theories of superconductivity, by Heisenberg (1947) and by Born & Cheng (1948), and we shall examine the relevant portions of these theories to see whether they may be regarded as essentially two-fluid theories. Both these theories have been subjects of considerable controversy, and it is not proposed here to add to the criticisms which have been levelled against them, but rather to examine their implications with regard to high-frequency effects in an endeavour to subject them to experimental test. In the theory of Born & Cheng it appears that the supercurrent is to be regarded as arising from the asymmetric orientation of filled and unfilled corners of the first Brillouin zone that is not completely occupied. Presumably those electrons which occupy the next higher zone will play no part in the process which, it is claimed, causes the corners of the first zone to be unequally filled, and under the influence of an electric field will therefore behave in the same way as the electrons in a normal metal. If this is a just interpretation of the model it is without question a two-fluid model of the type discussed, but confirmation of this view must await a more detailed statement of the theory.

A great deal, on the other hand, has been written about the theory of Heisenberg, so that it is easier to discuss its relation to the two-fluid model, which, at first sight, appears to be slight on account of the strong interaction between the condensed and uncondensed electrons. A concise account of the theory, together with references to fuller treatments, is to be found in Heisenberg's lecture (1949) and we shall not recapitulate what is so clearly expounded there. The treatment of the electrodynamics of this model is carried out by Heisenberg first on the assumption that mean free path effects may be neglected, and is then modified by the use of the ineffectiveness concept to take account of the long free path compared to the penetration depth. Here we shall depart from this method of approach and attempt to formulate the problem a little more precisely. An electric field acting on the electron assembly, of which a small fraction is composed of a condensed phase, the 'electron-lattices', moving with the Fermi velocity, transfers momentum to all the electrons and sets up a current in the metal. The effect of the field on the electron lattices is, however, small in comparison with the effect which arises from interactions between the lattices and the uncondensed, 'normal', electrons, which collide with the lattices and, if they have had imparted to them any drift momentum by the applied field, transfer a fraction of this momentum to the lattices. According to

Heisenberg this momentum transferred to the lattices, which is manifested by a change in their directions of motion, is not dissipated by collision with the metallic ions and constitutes the supercurrent. This part of the process may be formally described by means of a relaxation time τ_e , and the relation between supercurrent and normal current may be written thus:

$$\dot{J}_s = J_n/\tau_e,$$

or, if we consider a periodic process of angular frequency ω ,

$$i\omega\tau_e J_s = J_n. \quad (3)$$

Turning now to the behaviour of the normal electrons, we must take account of the fact that while in a normal metal they can lose momentum only by interaction with the ionic lattice, with a relaxation time τ_n , in the superconductor they may also lose momentum to the electron-lattices, with relaxation time τ_e . The resultant relaxation time for the normal electrons, τ' , is thus given by the equation

$$\frac{1}{\tau'} = \frac{1}{\tau_e} + \frac{1}{\tau_n}, \quad (4)$$

and the effective mean free path of the normal electrons, l' , may be defined in the usual way as $v\tau'$, where v is the Fermi velocity. Apart from this modification of the mean free path the behaviour of the normal electrons is unaltered, so that the relation between J_n and E may be written immediately in the form given by Reuter & Sondheimer (1948):

$$J_n = \frac{3\sigma'}{4l'} \int_{-\infty}^{\infty} k\left(\frac{x-t}{l'}\right) E(t) dt, \quad (5)$$

in which σ'/l' is the same as σ/l for the normal metal, i.e. ne^2/mv , and

$$k(u) = \int_1^{\infty} \left(\frac{1}{s} - \frac{1}{s^3}\right) e^{-s|u|} ds.$$

Since we are interested only in the form of the equations we have taken for simplicity a value unity for p , the surface reflexion coefficient. From equations (3) and (5) we may write down the relation between J_s and E

$$J_s = \frac{3\sigma'}{4i\omega\tau_e l'} \int_{-\infty}^{\infty} k\left(\frac{x-t}{l'}\right) E(t) dt. \quad (6)$$

Equations (5) and (6) show that although the supercurrent arises mainly from interaction between the normal electrons and the electron lattices, the two contributions to the total current may nevertheless be written independently. Heisenberg's model thus is to be regarded as a two-fluid model in the sense in which we have used the term. In its general form it is a model which is defined by more than two parameters, but under certain conditions all but two become unimportant, as we shall now show.

If the total current be written as the sum of J_n and J_s an equation for E may be derived with the help of Maxwell's equations

$$\frac{\partial^2 E}{\partial x^2} = \frac{3\pi\sigma'}{l'\tau_e} (1 + i\omega\tau_e) \int_{-\infty}^{\infty} k\left(\frac{x-t}{l'}\right) E(t) dt. \quad (7)$$

This equation has the same form as Reuter & Sondheimer's equation for the normal metal, with l replaced by l' and σ by $\sigma'(1 + 1/i\omega\tau_e)$. If then l' is much greater than the penetration depth λ , a condition which seems likely to hold in a real superconductor, the surface impedance Z may be written in the asymptotic form derived by Reuter & Sondheimer, with appropriate modifications,

$$Z = 3 \cdot 12 i \omega \left(\frac{l' \tau_e}{\sigma'} \right)^{\frac{1}{2}} (1 + i \omega \tau_e)^{-\frac{1}{2}}. \quad (8)$$

We may note that if the frequency is low, so that $\omega \tau_e \ll 1$, equation (7) becomes purely real:

$$\frac{\partial^2 E}{\partial x^2} = \frac{3 \pi \sigma'}{l' \tau_e} \int_{-\infty}^{\infty} k \left(\frac{x-t}{l'} \right) E(t) dt, \quad (9)$$

and the surface impedance correspondingly becomes purely reactive and proportional to ω :

$$Z_0 = 3 \cdot 12 i \omega \left(\frac{l' \tau_e}{\sigma'} \right)^{\frac{1}{2}}.$$

Thus the surface inductance, $Z_0/i\omega$, is independent of frequency, for not too high a frequency, and the corresponding penetration depth, $\lambda (= Z_0/4\pi i\omega)$, is also independent of frequency. Equation (9) is interesting as an example of an equation which reproduces the experimental requirement of a frequency-independent penetration depth, and yet is not a point-relationship like the London equation. If we substitute λ into equation (8), the surface impedance takes the simple form

$$Z = 4 \pi i \omega \lambda (1 + i \omega \tau_e)^{-\frac{1}{2}}. \quad (10)$$

Provided, then, that the mean free path is sufficiently long, Z is determined by only two parameters (apart from ω), λ and τ_e , and Heisenberg's model is reduced to a two-parameter two-fluid model.

In view of the preceding discussion there is some relevance in discussing the experimental results obtained at high frequencies in terms of the two-parameter two-fluid model, and the first stage in the discussion will be to make a dimensional analysis of the model. For this purpose it is necessary to choose two suitable parameters, of which one is conveniently taken to be the penetration depth λ . It is important to note that in making this choice we do not commit ourselves to the London theory, since λ is from an experimental point of view related to the total amount of field penetration into the surface, regardless of whether or not this penetration follows an exponential law. The choice of λ as an appropriate parameter for our purpose depends solely on the assumption that the equation relating J_s and E , whether it be equation (1) or (6) or some other, is such that there is only one combination of the quantities defining the state of the metal which is relevant to the solution of the electrodynamic problem. This condition is obviously satisfied by equation (1), and we have shown it to be satisfied also by equation (6) if l' is sufficiently great. For the other parameter we have a wide selection available, and in the following analysis we shall use several different parameters as convenient. All that is necessary is that it shall in some way define the behaviour of the normal component of the current, and in addition it is convenient that it should have the same dimen-

sions as λ . The obvious choice is what may be called 'the resistive skin depth of the normal electrons' δ'_r ; by this is meant, not that the normal electrons may under any circumstances be imagined as acting alone, the supercurrent being somehow inhibited, but that if the equation connecting J_n and E be used alone as if it referred to the total current, then δ'_r is the skin depth which the metal would exhibit. But δ'_r is only one of many possible choices, and any combination of δ'_r and λ which has the dimensions of length may equally well be taken.

THE SURFACE REACTANCE

It was seen in III how the presence of normal electrons might cause the inductive skin depth at 25 cm. to differ from the low-frequency penetration depth, especially near the transition temperature, and we shall now examine this effect a little more closely. In particular, the experimental results at 3 cm. presented in IV show that at the higher frequency the departure from the simple law which holds at low frequencies, $\lambda = \lambda_0(1 - t^4)^{-\frac{1}{2}}$, is more marked, and we must inquire whether the whole of this departure may be ascribed to the resistive mechanism, or whether there is in addition a dispersion of λ beginning to become apparent. We shall begin by assuming that there is no such dispersion and see what results may be deduced by dimensional means.

Let us define, as in III, a quantity x as the ratio, λ'/λ , of the observed inductive skin depth to the penetration depth. Then x is clearly a measure of the relative importance of the two groups of electrons, and will be some function of a dimensionless ratio of the relevant parameters describing the two-fluid model. For one of these we take λ and for the other δ_s , defined by the equation

$$\delta_s = R/2\pi\omega = r\delta_r, \quad (11)$$

where R is the surface resistance of the superconductor at the temperature considered, r is the ratio, R/R_n , of this resistance to the resistance of the normal metal just above the transition temperature, and δ_r is the resistive skin depth of the normal metal. Since δ_s is determined by both components of the electron assembly, it is a suitable parameter in accordance with the principles discussed above. We may therefore write

$$x = F(\delta_s/\lambda),$$

or, since x is itself dimensionless,

$$x = F'(\delta_s/x\lambda) = F'(\delta_s/\lambda'). \quad (12)$$

In this way x is seen to depend on measurable quantities, since r , δ_r and λ' are all found experimentally. We may note that at very low temperatures, when δ_s tends to zero and the normal electrons cease to play an important part, x may be expected to take the value unity, while at the transition temperature x equals zero, since λ' becomes equal to δ_i and λ becomes infinite. Thus x vanishes when the argument δ_s/λ' equals δ_r/δ_i , which from the definitions of the skin depths is equal to the ratio $2R/X$ for the normal metal. As a matter of interest the function F' has been computed for three models, H. London's (1940) which ascribes the classical conductivity

to the normal electrons, Sondheimer's which takes account of their anomalous behaviour, and Heisenberg's (from equation (10)), and the results are shown in figure 1. The different limiting values of δ_s/λ' are a consequence of the fact that for a classical metal $X/R = 1$, while for the extreme anomalous metal $X/R = \sqrt{3}$.

In order to display the experimental results in this way it is necessary to know both λ and λ' so that x may be calculated. For this purpose we shall assume that λ varies as $(1 - t^4)^{-\frac{1}{2}}$, in accordance with the results at low frequencies and with the behaviour of λ' at such low temperatures that λ and λ' should be equal. With this

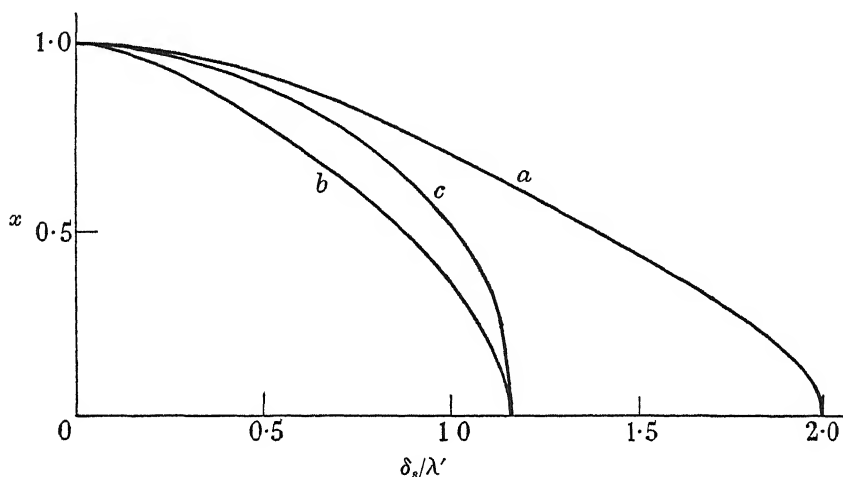


FIGURE 1. Influence of normal electrons on inductive skin depth, calculated for different models. *a*, H. London model, with or without ineffectiveness concept; *b*, Sondheimer model; *c*, Heisenberg model.

assumption x may be determined from the experimental measurements of reactance, while δ_s is calculated from the resistance measurements. Figure 2 shows the variation of x with δ_s/λ' for a 25 cm. specimen and for the two extreme 3 cm. specimens; the curve calculated from Sondheimer's theory is given for comparison. It is not feasible to extend the experimental curves to greater values of δ_s/λ' , since these correspond to temperatures very close to the transition temperature where the slight heating of the specimen by the measuring current is sufficient to render its precise temperature uncertain. It is clear from the curves that there is not very good agreement between experiment and the dimensional theory which predicts that all the curves should be the same. There are several hypotheses which might account for the discrepancies. For instance, the shape of the curves is very sensitive to experimental error, and particularly to a misjudgement of the slope at low temperatures of the lines drawn in figure 5 of IV, since this affects the initial slope of the curves in figure 2; nevertheless, it does not seem possible to account for all the discrepancy in this way. Again, the assumption of a simple law for the variation of λ with temperature may be in error, though this does not appear likely in view of the results obtained at 25 cm. and shown in figure 7 of IV, and the results by Laurmann & Shoenberg (1949) at low frequencies. There are left then the possibilities either that there is a new phenomenon, the dispersion of λ becoming apparent at the higher frequencies,

or that the two-parameter two-fluid model on which the theory is based is not an adequate description of a superconductor. In view of the evidence which may be obtained on these points from further analysis of the data we shall discuss them no further at the moment, but proceed to an analysis of the anisotropy of the surface reactance and its implications, contenting ourselves with the observation that at any rate a substantial proportion of the difference between λ' and λ may be ascribed to the influence of the mechanism which is responsible for the high-frequency resistance.

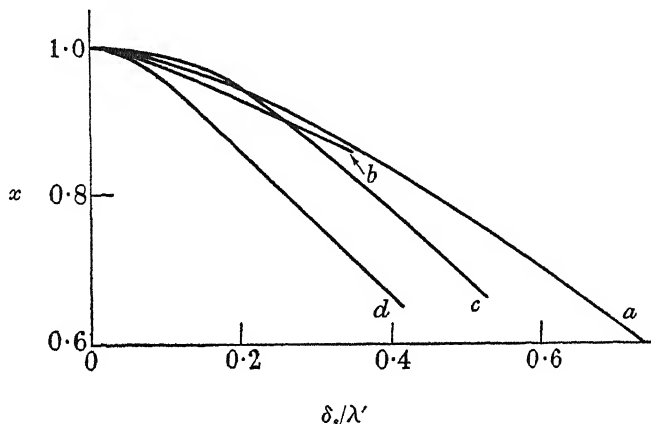


FIGURE 2. Influence of normal electrons on inductive skin depth found experimentally. *a*, theoretical curve for Sondheimer model, *b*, experimental curve at 25 cm.; *c*, experimental curve at 3.2 cm., $\theta = 87^\circ$; *d*, experimental curve at 3.2 cm., $\theta = 62.5^\circ$.

If it be assumed that the whole of the difference between λ and λ' is caused by the resistive mechanism the behaviour of x makes it clear that at the lower temperatures the value of λ' is the same as the value of λ measured with low frequencies or by static means. Since, therefore, for small enough t , $\lambda'(t) - \lambda'(0)$ obeys the same law at all crystal orientations, apart from a scaling factor, it must follow that $\lambda(t) - \lambda(0)$ also obeys the same law, in accordance with the observations of Laurmann & Shoenberg. Although absolute values of λ are not determined, but only changes with temperature, it seems a natural consequence of this result that λ itself obeys the same law, apart from a scaling factor, at all orientations, and that the quantity g plotted in figure 6 of IV is a measure of the value, λ_0 , of λ at 0°K . In fact, as was pointed out in IV, λ_0 is given by $14g$. Figure 6 may then be taken to represent the anisotropy of λ_0 . If this interpretation of the results is correct, it casts some doubt on the validity of the phenomenological theory of London & London. For their equation connecting \mathbf{J} and \mathbf{H} , $\text{curl } \Lambda \mathbf{J} + \mathbf{H} = 0$, extended in the natural manner to anisotropic crystals (Ginsburg 1944; v. Laue 1948) by treating Λ as a second-order tensor, implies that Λ , and hence λ^2 , may be represented by an ellipsoid. In a uniaxial crystal like tin the ellipsoid would be one of revolution about the tetrad axis, and hence λ^2 should be a linear function of $\cos^2 \theta$, in contradiction to our interpretation of the experimental results. This difficulty cannot be avoided by the supposition that λ_0 is not in fact proportional to g , but that λ_0^2 rather is a linear function of $\cos^2 \theta$, since at higher temperatures this linear relation would break down, the increment of λ being governed by g . It is possible, therefore, that the London equation must be replaced by some other

equation, such as equation (6), which is not subject to the restrictions of a point-relationship, and which may be consistent with the complicated anisotropy found experimentally.

On the other hand, it may be that the London equations are sound at low frequencies, but that at the high frequencies used in these experiments they break down as a result of dispersion or coupling between the two groups of electrons. Attractive though this alternative hypothesis may be, it must be admitted that the experimental evidence does not favour it. For the measured variation of λ with t follows at the lowest temperatures the same law as at lower frequencies, and moreover the measured rate of change of λ is in good agreement with the observations of Laurmann and Shoenberg for those crystal orientations which have been studied by the latter. It would be extraordinary if the new effect were to give rise to important modifications in the absolute values of λ while at the same time the rate of change were left unaltered. In addition, the agreement between experimental and theoretical values of δ_i/δ_r noted in IV suggests that the method used for estimating λ_0 for this purpose was sound, and that there is consequently no variation of λ_0 with frequency. There have been unfortunately insufficient measurements on the anisotropy of λ at low frequencies to decide whether this anomalous anisotropy also occurs there,* and a full understanding of these difficulties must therefore await the necessary experiments, which will be of considerable difficulty with known techniques. All that can usefully be said at the moment is that the weight of evidence supports the view that λ_0 is proportional to the measured quantity g , and that the anisotropy exhibited by λ is not in accordance with the London theory.

THE SURFACE RESISTANCE. DIMENSIONAL ANALYSIS

The analysis carried out in III of the resistance measurements at 25 cm. made it clear that the two-fluid model used there gave at least a qualitative explanation of the results. It would be possible to extend the same analysis to the present measurements, but it is more satisfactory to make a more general analysis by dimensional means, which will cover as a special case the model previously used. It is well at the very start to make clear which of the experimental results are to be used in the analysis, particularly as the anomalous anisotropy renders it unwise to draw deductions from some of the data. It was pointed out in IV that the surface impedance would be expected to vary at different azimuths of the same specimen, in spite of the fact that the direction of current flow is the same, since at different azimuths different groups of normal electrons will have their directions of motion lying in the surface and will make the most important contribution to the resistance. For this reason all the values of resistance probably represent some sort of average, and very little of quantitative importance may be deduced from the range of variation of, for instance, the scale factor G_2 . On the other hand, certain of the results are probably not the consequence of the averaging process, but represent the local behaviour of

* The measurements by Laurmann & Shoenberg (1949) on mercury suggest that for this metal at any rate the anisotropy is not in conflict with the London theory, but the precision of measurement is not quite high enough to make the matter absolutely certain.

every point on the surface. The results in this category which we shall use are the following:

(a) The penetration depth λ is the same function of temperature t for all orientations of the surface with respect to the crystal axis, apart from an orientation-dependent scale factor, g . The evidence for this point has already been sufficiently discussed above.

(b) The resistance in the transition region may be expressed by equation (3) of IV,

$$\Delta T = \text{constant} + G_1 \psi(r).$$

(c) The resistance in the low-temperature region may be expressed by equation (5) of IV,

$$r = G_2 f(t).$$

(d) G_1 and G_2 are proportional to one another.

We may confidently assume that these results, though deduced from measurements in which a certain amount of averaging occurs, are characteristic of local behaviour at all points of the specimen, or, in other words, of the behaviour of plane surfaces.

To carry out the dimensional analysis we shall introduce as parameters describing the two-fluid model the penetration depth, λ , and the 'skin depth of the normal electrons', δ'_r , whose significance was expounded above. In considering the quantities on which δ'_r depends it is convenient to recall the formula for δ_r in the normal metal under extreme anomalous conditions,

$$\delta_r = a \left(\frac{\omega \sigma}{l} \right)^{-\frac{1}{2}},$$

where a is a numerical constant. This expression ensures that δ_r shall be independent of the mean free path of the electrons, as found in practice. Now in the measurements on superconductors there is no indication of a dependence of the resistance on the mean free path, and it is therefore plausible to take the same form for δ'_r ,

$$\delta'_r = a \left(\frac{f_n \omega \sigma}{l} \right)^{-\frac{1}{2}} = f_n^{-\frac{1}{2}} \delta_r. \quad (13)$$

The new quantity, f_n , introduced here may be interpreted on the basis of the model used in III as the fraction of electrons remaining in the normal state, or alternatively it may be regarded simply as a temperature-dependent parameter introduced to allow for the difference between δ'_r and δ_r . It is to be expected, however, that f_n will be closely related to the fraction of normal electrons which contribute to the specific heat in the theory of Gorter & Casimir (1934); it must take the value unity at the transition temperature, and, since the transition is of the second order, will probably vary continuously, rather than abruptly, as the temperature is lowered.

Now, it is very important for the following analysis to decide whether f_n is to be regarded as a function of temperature alone, or of orientation as well. There is no reason to suppose *a priori* that f_n will be independent of orientation, since as the superconducting transition proceeds it is likely that different parts of the Fermi surface will be differently affected. Nevertheless, it appears from the experimental results that f_n may be regarded as independent of orientation. In the first place we have found that the r against t curve has the same functional form for all orientations,

and for values of t less than 0.86; if, therefore, the form of the λ against t curve is independent of orientation, so also must be the variation of δ'_r , and consequently f_n , apart possibly from a scale factor. That there should be such an orientation-dependent scale factor is improbable in view of the fact that f_n must attain the value unity for all orientations at the transition temperature. In the second place the linear correlation between G_1 and G_2 could hardly arise if f_n were dependent on orientation, since only G_2 is to any extent influenced by the value of f_n , G_1 being determined by the variation of r so close to the transition temperature that f_n may be taken as unity. There is thus good reason for assuming f_n to be a function of temperature only, and we shall proceed with the analysis on this assumption.

By analogy with the definition of δ'_r (equation (13)) we may define a corresponding resistance $R'_n (= 2\pi\omega\delta'_r)$ as the surface resistance of a metal in which the current and field are related by the same equation as relates J_n to E in the two-fluid model. Thus

$$R'_n = f_n^{-\frac{1}{2}} R_n. \quad (14)$$

Then if R is the resistance of the superconductor for any temperature and orientation, we may set up a dimensional equation as follows.

$$\frac{R}{R'_n} = \Phi\left(\frac{\lambda}{\delta'_r}\right).$$

It is important to note that ω cannot be introduced explicitly into this equation without making it dimensionally inhomogeneous, but that it occurs implicitly in δ'_r , as also does f_n . There is no reason to introduce f_n explicitly since R is derived for any particular model from a solution of the equation for the total current $J_n + J_s$, and R'_n from a solution of the equation for J_n alone at the same temperature; these equations contain as parameters only λ and δ'_r . If, however, we wish to find the ratio R/R_n we must introduce f_n , since R and R_n are measured at different temperatures and depend upon different values of f_n . Thus, from equation (14),

$$r = \frac{R}{R_n} = f_n^{-\frac{1}{2}} \Phi\left(\frac{2\pi\omega f_n^{\frac{1}{2}} \lambda}{R_n}\right). \quad (15)$$

This is the fundamental equation which we may now apply to the experimental results. First let us consider the transition region, for which f_n may be taken as unity. We may also conveniently use Shoenberg's (1940) expression for the penetration depth near the transition temperature (which is consistent with the law,

$$\lambda = \lambda_0 (1 - t^4)^{-\frac{1}{2}},$$

when $1 - t \ll 1$),

$$\lambda \propto \lambda_0 (\Delta T)^{-\frac{1}{2}},$$

writing ΔT for $(1 - t) T_c$. Equation (15) then takes the form

$$r = \Phi\left(\frac{2\pi\omega\lambda_0}{R_n(\Delta T)^{\frac{1}{2}}}\right).$$

This equation implies a similarity of the initial portions of the curves for different orientations, as is observed in practice, and enables the scale factor G_1 to be related

to other measurable quantities, since G_1 is proportional to the value of ΔT at which r takes some standard value. Hence

$$\frac{\omega \lambda_0}{R_n G_1^{\frac{1}{2}}} = \text{constant},$$

and, since $\lambda_0 \propto g$,

$$G_1 \propto \left(\frac{\omega g}{R_n} \right)^2. \quad (16)$$

For different orientations, but constant frequency, we should thus expect g to vary as $R_n G_1^{\frac{1}{2}}$. Now, on account of the averaging effect produced by the use of cylindrical specimens, it is not possible to compare this expression exactly with the experimental results. It is worth while, however, to plot $R_n G_1^{\frac{1}{2}}$ against θ to see that there is no marked difference between this curve and the curve for g (figure 6 of IV), and this has been done in figure 3. It is clear that the curves are of the same general shape, although the range of variation is not so great for $R_n G_1^{\frac{1}{2}}$ as for g . In view of the uncertainty about the importance of the averaging effect this comparison may be regarded as satisfactory.

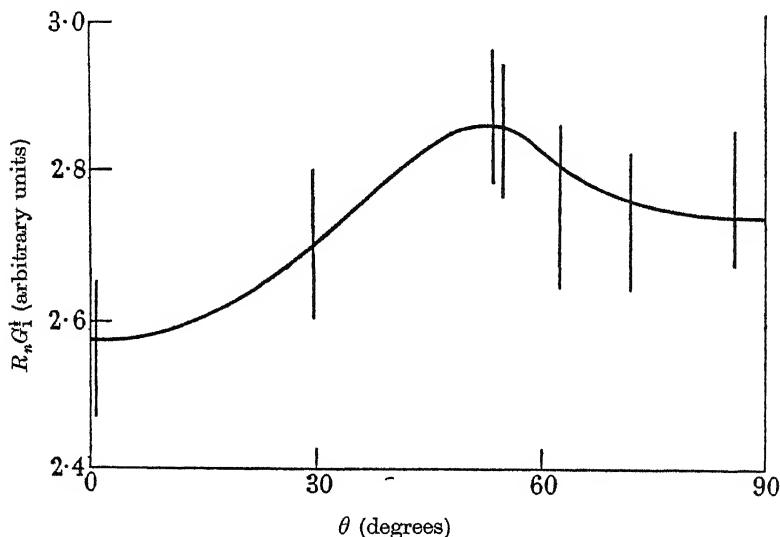


FIGURE 3. Anisotropy of $R_n G_1^{\frac{1}{2}}$.

A further comparison of equation (16) with experiment may be obtained from the variation with frequency of the initial slope. Since R_n varies as ω^3 we should expect G_1 also to vary as ω^3 , so that it should increase by a factor 3.95 between 25 and 3.2 cm. Unfortunately, the measurements at 25 cm. are not very reliable on account of the inevitable rounding off produced by the spread of the transition even under stationary conditions; moreover, no study was made of the anisotropy and the orientations of the specimens were not determined. The values of G_1 at 3.2 cm. lay between 1.75 and 3.58, so that at 25 cm. one would expect values between 0.44 and 0.91; for one specimen at 25 cm. G_1 was about 0.70, so that there is no evidence of any disagreement between theory and experiment.

In order to analyze the behaviour of the resistance in the lower temperature range, where r may be expressed as $G_2 f(t)$ (equation (5) of IV), we must return to equation

(15), not now neglecting f_n . If for λ we substitute $z\lambda_0$, where $z = (1 - t^4)^{-\frac{1}{2}}$, we have that

$$r = f_n^{-1} \Phi(\omega z \lambda_0 f_n^{\frac{1}{2}} / R_n). \quad (17)$$

Since the value of r at some standard temperature is a measure of G_2 , and since we assume f_n to depend only on temperature, we may write

$$G_2 \propto \Phi(b\omega\lambda_0/R_n), \quad (18)$$

where b is a constant, being the value of $zf_n^{\frac{1}{2}}$ at the standard temperature.

Now, experimentally we have found that for different orientations G_1 and G_2 are proportional, and we may use this result to determine the form of the function Φ . If we compare equations (16) and (18), remembering that $\lambda_0 \propto g$, we see that Φ is a quadratic function,

$$\Phi(u) \propto u^2.$$

The range of u over which this expression holds is governed by the range over which r is proportional to $f(t)$, since the standard temperature used to define G_2 may be any within this range. Roughly, then, Φ is quadratic provided that r is less than 7 %, and within this range we may rewrite equation (17) in the form

$$r \propto \omega^2 z^2 \lambda_0^2 f_n^{\frac{1}{2}} / R_n^2. \quad (19)$$

From this equation the frequency variation of r in the lower temperature range may be predicted. For any given temperature and orientation $z^2 \lambda_0^2 f_n^{\frac{1}{2}}$ is constant, so that r should be proportional to ω^2 , since $R_n \propto \omega^{\frac{1}{2}}$. Hence if r is proportional to $f(t)$ at 3.2 cm. it should also be proportional to $f(t)$ at 25 cm., the constant of proportionality being less by a factor 3.95. It is clear from figure 11 of IV that the first part of this prediction is well verified, but the ratio of the slopes is in not quite so good agreement. At 3.2 cm. G_2 varies between 9 and 20, so that at 25 cm. the variation should be between 2.3 and 5.1; for the specimen of figure 11 the value of G_2 is 2.3, just within the limits of variation, but it is likely that this value is appropriate to an orientation which leads to a rather high value of G_2 , since the average value for the measurements of I was only 1.95, outside the limits predicted by the theory.

THE SURFACE RESISTANCE: ELECTRODYNAMICAL CONSIDERATIONS

Thus far, apart from the small discrepancy just noted, the two-fluid model appears to give a consistent formulation of the observed behaviour, and it would be tempting to claim that the model is justified were it not for the following considerations of what is implied by equation (19). We have already noted that r should vary as $\omega^{\frac{1}{2}}$, and in addition equation (19) shows that it should vary as $f_n^{\frac{1}{2}}$. Now the essential characteristic of the two-fluid model lies in the possibility of dividing the current into two independent parts, normal and superconducting, and it is clear that at the lowest temperatures, where r is very small, the greater part of the current is carried by the superelectrons. It is to be expected that under these conditions the superelectrons will be largely responsible for determining the configuration of electric and magnetic fields in the skin layer, and this expectation is substantiated by the behaviour of x , which shows that the normal electrons do not influence the surface reactance if r is

very small. Provided, then, that λ_0 is independent of frequency the magnitude and configuration of the magnetic field should be independent of frequency for a given total current, and consequently the electric field, which may be regarded as being induced by the oscillatory magnetic field, will have a fixed configuration but a magnitude proportional to the frequency. The normal electrons moving in this field will carry a current proportional to the field, and give rise to a power dissipation, and hence a resistance, proportional to the square of the field; the surface resistance, R , should thus vary as the square of the frequency, and, since $R_n \propto \omega^{\frac{1}{2}}$, r should vary as $\omega^{\frac{1}{2}}$. In addition, R should be proportional to the number of normal electrons, since they do not appreciably affect the field distribution and may therefore be regarded as acting independently.

These requirements would be satisfied if Φ were a quartic rather than a quadratic function, for then equation (17) would become

$$r \propto \omega^4 z^4 \lambda_0^4 f_n / R_n^4,$$

so that r would vary as $f_n \omega^{\frac{1}{2}}$, since $R_n \propto \omega^{\frac{1}{2}}$. But this quartic form of Φ would imply that G_2 should vary as G_1^2 , and inspection of figure 12 of IV shows this form of variation to be quite inadmissible. The experimental behaviour is thus in conflict with the general considerations of the last paragraph, considerations which may readily be justified by a more detailed examination of any particular two-fluid model. For example, the expression given in equation (10) for the surface impedance according to Heisenberg's theory may be expanded in series and separated into real and imaginary parts

$$R = 4\pi\omega\lambda(\frac{1}{3}\omega\tau_e - \frac{1}{81}\omega^3\tau_e^3 + \dots),$$

and

$$X = 4\pi\omega\lambda(1 - \frac{2}{3}\omega^2\tau_e^2 + \dots).$$

If then $\omega\tau_e$ is so small that X is not appreciably different from $4\pi\omega\lambda$, it is permissible to neglect all but the first term in the expansion of R , which then becomes proportional to ω^2 , in agreement with our general considerations.

Additional evidence that equation (19) cannot be correct may be found in the temperature variation of f_n which may be deduced therefrom. Substituting the known temperature variation of r and z into the equation, we find for the variation of f_n

$$f_n \propto t^{12}(1+t^2)^{-3}. \quad (20)$$

Now although f_n may not be exactly the same as the fraction of normal electrons which occurs in the thermodynamical theory of Gorter & Casimir (1934), we should expect the two quantities to be fairly closely related; that is, f_n is unlikely to vary much more rapidly than t^4 . However, when t changes from 0.4 to 0.8, the temperature range over which equation (20) should be valid, f_n changes by a factor 1450, while t^4 changes by only 16. This enormous discrepancy makes it clear that equation (20) cannot represent the true variation of f_n , and this conclusion, taken in conjunction with the considerations of the frequency variation of r , points undoubtedly to the incorrectness of equation (19) and the reasoning leading up to it. It might be argued that the error lies not so much in the two-fluid hypothesis as in the two-parameter hypothesis, and certainly if more parameters are introduced the whole dimensional

argument needs revision. Nevertheless there seems to be no reason why the arguments of this section should not still lead to an expected variation of R as the square of the frequency, and therefore to a contradiction with the experimental results.

DISCUSSION

The most positive conclusion that can be drawn from this analysis is that the experimental results constitute a refutation of the two-fluid model as formulated here, and therefore of the theories, such as Heisenberg's, which have been shown to be particular examples of this model. Less certain is the doubt cast on the phenomenological theory of London & London by the anomalous anisotropy of the reactance. It is not easy to see in what direction one should proceed to try to formulate electrodynamical equations for a superconductor which leads to predictions in agreement with experiment. If the two-fluid model is to be retained it seems essential that the penetration depth shall decrease with increasing frequency so that the resistance may increase less rapidly than as the square of the frequency. This can certainly be achieved by introducing linear coupling coefficients between the two groups of electrons which depend, for example, on \vec{J}_n or \vec{J}_s , but, as well as being a highly artificial contrivance, this type of modification is unlikely to give rise to a decrease in the penetration depth at high frequencies while leaving unaltered the temperature variation of the penetration depth, as is required by the experimental results. It is much more probable that the two-fluid model will have to be abandoned, and that the electrodynamics will be capable of formulation, if at all, only in terms of the total current J rather than two separate contributions J_n and J_s . At the same time it must not be forgotten that the two-fluid model gives a satisfactory correlation of the thermodynamical properties and the temperature variation of penetration depth, and in addition provides a model of the second-order transition between the superconducting and normal states. No alternative model has yet been proposed which satisfies all these requirements, and there is therefore at present little likelihood of finding an interpretation of these experiments.

I should like to thank Dr D. Shoenberg for his continued interest in this problem and for much helpful advice. I have also benefited greatly from discussions of this and related topics in superconductivity, particularly with Professor W. Heisenberg and Dr E. H. Sondheimer.

REFERENCES

- Becker, R., Heller, G. & Sauter, F. 1933 *Z. Phys.* **85**, 772.
 Born, M. & Cheng, K. C. 1948 *Nature*, **161**, 968 and 1017.
 Daunt, J. G., Miller, A. R., Pippard, A. B. & Shoenberg, D. 1948 *Phys. Rev.* **74**, 842.
 Ginsburg, V. L. 1944 *J. Phys. U.S.S.R.* **8**, 148.
 Gorter, C. J. & Casimir, H. B. G. 1934 *Z. tech. phys.* **15**, 539.
 Heisenberg, W. 1947 *Z. Naturforschung*, **2a**, 185.
 Heisenberg, W. 1949 *Two lectures* (Cambridge University Press).
 v. Laue, M. 1948 *Ann. Phys., Lpz.*, **3**, 31.
 Laurmann, E. & Shoenberg, D. 1949 *Proc. Roy. Soc. A*, **198**, 560.

- London, F. & London, H. 1935 *Proc. Roy. Soc. A*, **149**, 71.
 London, H. 1934 *Nature*, **133**, 497.
 London, H. 1940 *Proc. Roy. Soc. A*, **176**, 522.
 Maxwell, E., Marcus, P. M. & Slater, J. C. 1949 *Phys. Rev.* **76**, 1332.
 Pippard, A. B. 1947 *Proc. Roy. Soc. A*, **191**, 370, 385 and 399 (parts I, II and III).
 Pippard, A. B. 1950 *Proc. Roy. Soc. A*, **203**, 98 (part IV).
 Reuter, G. E. H. & Sondheimer, E. H. 1948 *Proc. Roy. Soc. A*, **195**, 336.
 Shoenberg, D. 1940 *Proc. Roy. Soc. A*, **175**, 49.

Field variation of the superconducting penetration depth

BY A. B. PIPPARD

The Royal Society Mond Laboratory, University of Cambridge

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 3 May 1950)

The high-frequency technique developed previously has been used to investigate the dependence of the penetration depth in superconducting tin on the strength of the applied field. A steady magnetic field was applied transverse to a thin cylindrical specimen forming part of a 3 cm. resonator, any resulting change in penetration depth being revealed as a shift in the resonant frequency. The change was greatest close to the transition temperature (3.72° K), but even here it amounted to no more than 3 % at the critical field strength. As the temperature was lowered the effect became smaller at first, reaching a very low value at 3° K, and then increased once more to about 2 % at 1.7° K. The effect of a steady magnetic field on the high-frequency resistance was also studied, though not in great detail. Above 3° K the resistance decreases slightly until at a field strength rather greater than $\frac{1}{2}H_c$ a sudden increase announces the formation of the intermediate state. At lower temperatures the effect of the steady field is to increase the resistance slightly. It is suggested that the very small change in penetration depth even at the critical field strength is evidence of the existence of long-range order in the superconducting state over a distance of 10^{-4} cm. or more, and the very sharp resistance transition in pure superconductors is adduced as further evidence of this hypothesis.

INTRODUCTION

It has often been suggested that the depth of penetration, λ , of a magnetic field into a superconductor should be regarded not as independent of field strength at constant temperature but rather as increasing with the field. Thus H. London has pointed out (in a private communication) that the temperature variation of λ implies thermodynamically a field variation of entropy, corresponding to a modification of the electronic configuration on application of a field, which would be expected to produce some change in λ . A similar argument has been applied by Ginsburg (1947) to a model of a superconductor based on Landau's (1941) quantum hydrodynamics. For quite other reasons, arising out of Heisenberg's (1947) theory of superconductivity, Koppe (1949) has suggested a strong field-dependence of λ particularly at very low temperatures, and this idea has been considerably developed by von Laue (1949). Up to now, however, experimental information has been very scanty, being confined to the observation by Laurmann & Shoenberg (1949) that just below the transition temperature of mercury there was no change of λ greater than 10 % in fields up to 80 % of the critical field.

The microwave technique (Pippard 1950*a*, referred to in future as A) which was developed for studying the temperature variation of λ is perhaps the only one known at present of sufficient sensitivity to measure with any precision changes of only a few per cent, and it is fortunately very easily adapted for measurement in the presence of a steady magnetic field. With its aid determinations have been made at a number of temperatures of the variation of λ in tin resulting from the application of fields up to, and even exceeding, the critical field strength, H_c . The first part of this paper describes the technique and experimental results, together with a number of associated phenomena which appeared during the course of the work and which, not being directly relevant to the investigation, have not yet been examined in detail. The second part is concerned with a discussion of the results, particularly the fact that the variation is considerably smaller than that expected on the basis of the arguments mentioned above. It is suggested that the experiments indicate the existence of long-range order in the superconducting state, and confirmation of this view is shown to be provided by the sharpness of the resistance transition of a pure superconductor in the absence of a magnetic field.

EXPERIMENTAL METHOD

A full description of the resonance techniques for determining the surface resistance and reactance of a metal at wave-lengths around 3.2 cm. has been given in A, and the same method was used in the present work. The specimen consisted of a thin wire of tin cast in a thin-walled silica tube, the length being 1.4 cm., nearly one-half of a wave-length. It was supported by silica threads along the axis of a cylindrical brass can, forming with it a co-axial resonator. Measurement of the band width of the resonator gave a value for the surface resistance of the specimen (after correction for extraneous losses in the resonator and coupling circuits), and measurement of the change in resonant frequency on application of a steady field enabled the corresponding change in λ to be calculated.

A large Helmholtz pair of coils outside the cryostat provided a uniform magnetic field transverse to the specimen; care was taken to use only non-superconducting solders close to the specimen to avoid inhomogeneity of the field. The choice of a transverse rather than a longitudinal field was conditioned by a number of factors. The main advantage of a longitudinal field is that it provides a uniform field over the surface of the specimen, while a transverse field is distorted by the superconducting specimen so that the field at the surface varies from point to point between zero and twice the applied field. Apart from this, however, there are two reasons for favouring the transverse field: first, the currents induced by it in the specimen are in the same direction, parallel to the axis of the specimen, as the measuring high-frequency current, and secondly, on account of the delay in the transition of small specimens into the intermediate state until the applied transverse field is greater than $\frac{1}{2}H_c$, there is the possibility of measuring λ under such conditions that some regions of the specimen are subjected to fields greater than H_c . These advantages were felt to outweigh the disadvantages of a non-uniform field at the surface, especially as a correction for non-uniformity may easily be applied once experiment has revealed the law governing the variation of λ with field strength.

A number of preliminary experiments were performed to ascertain the general features of the effects due to the applied field and to determine the most suitable diameter for the specimen. The final measurements were made on two specimens of circular cross-section which were rather thinner than those used in the previous work, one having diameter 93μ and the other 58μ . No precautions were taken in casting the specimens to ensure that they were single crystals, but it is probable, judging from previous experience, that they consisted of not more than a few crystallites, and the sharpness of the transition into the intermediate state suggests that the thicker at any rate was a single crystal.

Measurements were made of the effect of a magnetic field on both the resistance and the reactance of the specimens. The variation of resistance with temperature was determined in the usual way from the width of the resonance curve, but the variation with field strength was studied by a different technique which was sufficiently accurate and much more rapid. The coupling between the specimen and the external circuits was fixed at a suitable value, and the frequency of the oscillator adjusted to the resonant frequency; the magnetic field was then raised by stages, and changes in the resonator output noted, the frequency being adjusted as the resonant frequency altered. The conversion of such readings into values of surface resistance presents no problem, so that it was possible in this manner to follow the changes in resistance as the specimen passed through the superconducting into the intermediate, and thence into the normal state. Similar measurements were made as the field was reduced once more to zero. After each run of this nature a large field, greater than the critical, was switched on and off again, in order to eliminate any normal inclusions which might have been frozen in during the slow reduction of field, although in fact there was never evidence of such inclusions, the specimens exhibiting an apparently perfect Meissner effect.

The changes in λ caused by the application of fields less than that required to establish the intermediate state were determined from the resulting change in resonant frequency, and the smallness of these changes enabled a very simple technique to be used. The oscillator frequency was adjusted so that the output of the resonator was three-quarters of the peak output (at this level the variation of output with frequency is most rapid), and observations were made of the change in output on application of the field. The measurements were repeated on the other side of the resonance peak so that allowances might be made for changes in the peak output and width of the resonance curve. From these observations the change in resonant frequency was determined relative to the band width of the resonator, already known from the resistance measurements. This method of measuring small changes in resonant frequency is very sensitive, especially at the lower temperatures where the band width is small, and over most of the range below 3.7°K it was possible to detect the change in frequency resulting from a change in λ of only $5 \times 10^{-9}\text{ cm.}$, about $\frac{1}{16}\%$ of λ itself. The factor relating frequency to penetration depth was calculated for each specimen from its dimensions, since the small diameter excluded its experimental determination as in A, the band width being too great for accurate measurement at temperatures high enough for the classical skin effect theory to apply. It is believed that the calculated conversion factors are not in error by more

than a few per cent, and in any case an error will affect only the absolute values of the changes in λ , not the relative values since the same conversion factor enters in the determination of λ by the method described in *A*.

EXPERIMENTAL RESULTS

(a) Resistance

The variation of the surface resistance with applied magnetic field is exemplified by the curve of figure 1, although at other temperatures and with the other specimen there were differences in detail. In the diagram the resistance R is expressed as a percentage of the resistance R_n of the normal metal at the same temperature. The resistance remains nearly constant until the field attains a value considerably greater

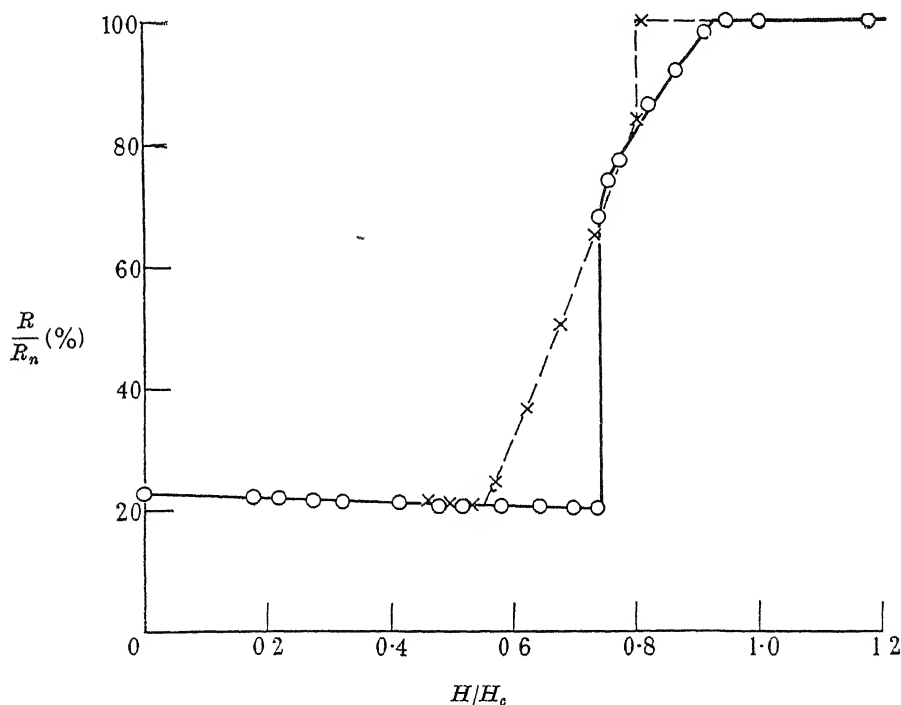


FIGURE 1. Field variation of high-frequency resistance. Thicker specimen at 3.546°K ; $H_c = 26.5 \text{ G}$; full curve, field increasing; broken curve, field decreasing.

than $\frac{1}{2}H_c$, when there is a sudden transition into the intermediate state. For the thicker specimen this transition was discontinuous except at temperatures just below T_c , but for the thinner specimen it was not discontinuous, but covered a small range of field strength, usually about $0.02H_c$. The fact that in the thinner specimen the transition occurred at fields about 0.6 rather than $0.74H_c$ suggests that the portion of the curve between 0.6 and $0.74H_c$ in figure 1 represents the metastable persistence of superconductivity, analogous to the superheating of a liquid. Between 0.74 and $0.94H_c$ the specimen is in the intermediate state, and the measured resistance is presumably an average of contributions from regions of the surface which are

either superconducting or normal. So far as can be seen from this curve the surface becomes wholly normal in a field less than H_c , but the precision of measurement is hardly high enough to preclude the possibility of slight traces of superconductivity persisting up to or beyond H_c . As the field is reduced there is at first a metastable persistence of the normal state ('supercooling'), and when the intermediate state is eventually established there is a steady transition into the superconducting state at $0.55H_c$. Apart from the possible superheating the behaviour parallels that observed by Désirant & Shoenberg (1948) and Andrew (1948) in their investigations of the magnetic behaviour and conductivity of thin wires in a transverse field.

An unexpected observation was that of a decrease of resistance in fields too small to cause a transition into the intermediate state. This effect was noticed in all the specimens, including three which were used for the preliminary investigation, but not at all temperatures. It does not begin until the temperature is below 3.63°K , the transition temperature being 3.72°K , but is already well marked at 3.546°K , the temperature at which the curve of figure 1 was measured. Here R/R_n falls from 22.8 % in zero field to 20.2 % at $0.74H_c$. At 3.0°K the drop was from 5.3 to 4.4 %, and below this temperature there is a tendency for the reverse effect to set in. Thus at 2.6°K there was an initial drop from 2.4 to 2.2 % at $0.3H_c$, followed by a rise to 2.4 % at $0.68H_c$, where the transition to the intermediate state occurred. At the lowest temperature, 1.7°K , there was no drop but a steady rise from 0.4 to 0.7 %. These results were obtained with the thicker specimen, the thinner showing similar effects but to a slightly smaller extent. No such behaviour is exhibited by the normal metal, so that the phenomenon must be regarded as characteristic of the superconductor. Since the drop is observed in fields as small as 10 G it is not likely that it can be due to the effect discovered by Macdonald (1949) in thin wires under the influence of a magnetic field. It is particularly difficult to understand how a drop in resistance can accompany an increase in penetration depth as is indicated by the reactance which will be treated in the next section.

Other points of interest in the behaviour of the resistance concern the metastable persistence of superconductivity and the sharp transition to the intermediate state. The sharpness is remarkable in view of the fact that the strength of the measuring field was by no means negligible. No equipment was available for measuring the power supplied to the resonator, and hence for calculating the oscillating field strength, but rough estimates were possible from a knowledge of the rated power output of the oscillator and of the order of sensitivity of the crystal detector. In this way it was estimated that the peak field strength at the centre of the specimen was about 3 G at T_c , increasing to 10 G at 1.7°K . This oscillating field did not prevent the specimens from exhibiting a sharp transition at T_c , nor did it tend to precipitate the intermediate state from the superheated superconducting state. The influence of the measuring field on the point at which the intermediate state was formed was studied at 1.7°K , where it was found that a threefold increase (from 10 to 30 G if the estimate is correct) did not destroy the abruptness of the transition, but merely caused it to take place when the applied field was smaller by 2 G. It appears therefore that a field oscillation at 10^4 Mc./sec. is too rapid to be followed by the superconductor, and that the relaxation frequency for the transition from the superconducting to the normal

state, if it is indeed definable, must lie between 10^4 Mc./sec. and the frequency of 20 Mc./sec. at which it seems the transition may still occur (Lasarev, Galkin & Khotkevich 1947).

The field strength at which the transition to the intermediate state occurs varies in a complicated way as the specimen is rotated about its axis, the direction of the field being fixed, as shown in figure 2, for which the points were taken with the thicker specimen at 3.33° K. It will be noted that for a given orientation the behaviour is fairly reproducible, and that the curve, apart from fine details, has the expected periodicity of 180° . The most plausible explanation is that there are small patches

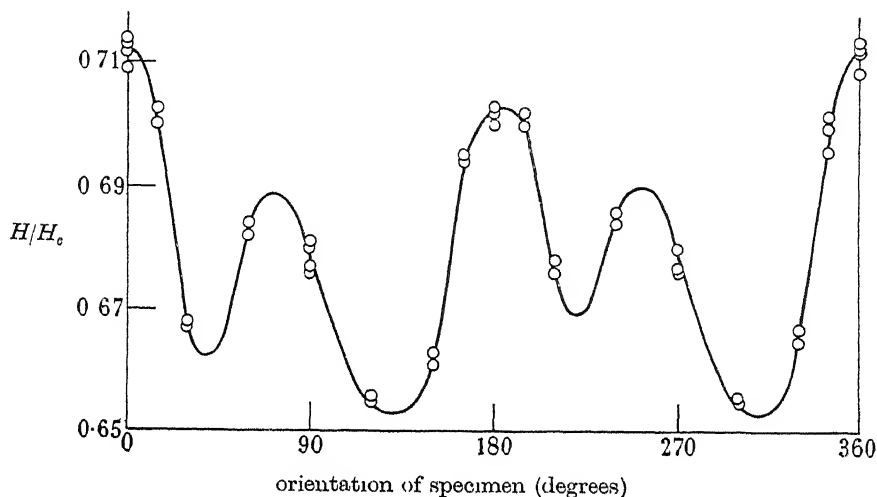


FIGURE 2. Variation with orientation of field at which intermediate state is formed. Thicker specimen at 3.330° K; $H_c = 59$ G.

on the surface which because of strain are favourable for the formation of normal nuclei, and that when the specimen is oriented so as to bring them into the region of high field strength they enable the intermediate state to be precipitated earlier. The experiment was repeated at 1.8° K, when it was found that for a given orientation the behaviour was much less reproducible, the transition occurring at any field strength within a range of as much as 10 %. Clearly no fundamental conclusions can be drawn from the actual value of the field strength at which the superconducting phase breaks down.

(b) Reactance

As explained above, measurements of the change in resonant frequency may be converted into changes in inductive skin depth, λ' , by means of a conversion factor depending on the dimensions of the resonator. The inductive skin depth is defined as $X/4\pi\omega$, where X is the surface reactance. At the lowest temperatures λ' is believed to be the same as λ , while nearer the transition temperature λ' is smaller than λ on account of the mechanism responsible for the high-frequency resistance (Pippard 1950*b*, referred to in future as B). Now the local field strength, H_l , at the surface of the specimen varies between 0 and $2H$, if H is the applied field, and the measured change of inductive skin depth, $\overline{\Delta\lambda'}$, is an average of local values of $\Delta\lambda'$ corresponding

to values of H_i within this range. Since the distribution of field around the specimen is known it is not difficult to calculate from any observed law of variation of $\overline{\Delta\lambda'}$ with H how $\Delta\lambda'$ varies with H_i .

Typical results for the variation of $\overline{\Delta\lambda'}$ with H^2 are shown in figure 3, from which it may be seen that the variation is very nearly quadratic in H except perhaps for

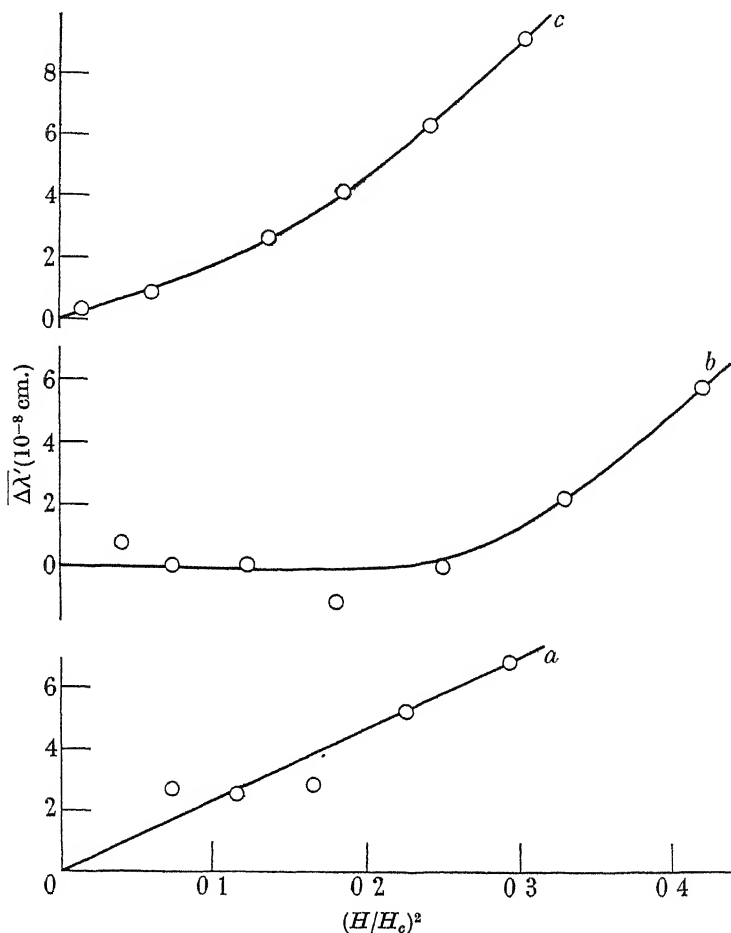


FIGURE 3. Variation of $\overline{\Delta\lambda'}$ with $(H/H_c)^2$. *a*, thinner specimen at 3.480° K, $H_c = 36.0$ G; *b*, thicker specimen at 3.036° K, $H_c = 97.5$ G; *c*, thinner specimen at 2.139° K, $H_c = 198$ G.

a short range above $\frac{1}{2}H_c$ before the intermediate state supervenes. A short calculation shows that over the quadratic range the curves may be converted to represent the variation of $\Delta\lambda'$ with $(H_i/H_c)^2$ by multiplying the abscissae by 4 and the ordinates by 2. The conversion of $\Delta\lambda'$, the change of inductive skin depth, into $\Delta\lambda$, the change of penetration depth at zero frequency, is not so simple. From the detailed discussion of this problem in B it may be seen that no serious error is likely to result from taking λ' and λ as equal below 3° K, but between this temperature and T_c the difference may be significant. The factor relating λ' to λ in zero field is known with reasonable accuracy from the previous work, but in the presence of a field the problem is complicated by the way in which the resistance decreases appreciably on application

of the field. It has been assumed here that the conversion factor is not affected by this decrease. If this assumption is incorrect, the most probable alternative is that the decrease in resistance causes λ' to be more nearly equal to λ ; thus even if λ were unchanged by the application of the field λ' would show an increase, and consequently the assumption made here is likely to lead to a value of $\Delta\lambda$ greater than the correct value. The values of $\Delta\lambda$ presented here are therefore probably maximum values, but this cannot be verified until a method has been devised for measuring changes of penetration depth at lower frequencies with a sensitivity comparable to the present method.

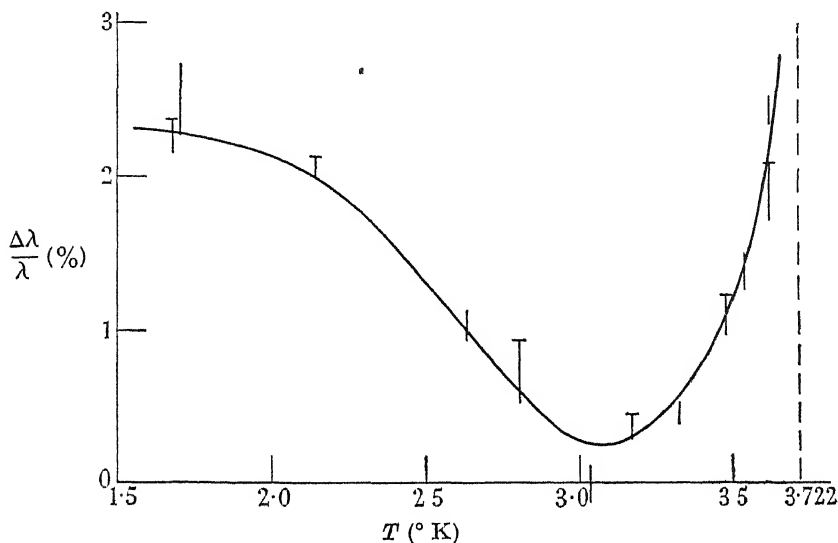


FIGURE 4. Percentage change in λ between zero and critical field, as a function of temperature. The length of each line represents the estimated probable error. The lines marked with a cross-stroke were obtained with the thinner specimen.

The final results of the measurements on the two specimens are exhibited in figure 4, which shows the percentage change in penetration depth between zero and the critical field as a function of temperature. It will be seen that the effect is very small at all temperatures, so that for most purposes λ may be considered as independent of H , and the superconductor treated as a linear system.

DISCUSSION

(a) Interpretation of experimental results

Apart from the smallness of the change in penetration depth, with which we shall deal mainly in this section, the most obvious point of interest is the shape of the curve in figure 4, especially the way in which, as the temperature is reduced, it falls to a very low value and then increases once more. As we shall see later, the effect would be expected to decrease steadily with lowering of the temperature, becoming zero at 0°K, and this expectation combined with the already discussed change in the behaviour of the resistance around 3.0°K suggests that there may be two mechanisms at work, dominant respectively at temperatures above and below 3.0°K. It is just

conceivable that the low temperature effect is the saturation of supercurrent discussed by Koppe (1949), but, apart altogether from doubts concerning the validity of the theory on which his work is based (London 1948; Pippard 1950*b*), the observed increase in penetration depth seems too small and too gradual in its inception to be thus explained. It is also unlikely to be due to the production of small regions of normal material on the surface, since the increase begins well before the field reaches the value $\frac{1}{2}H_c$ and the transition to the intermediate state is just as abrupt as at higher temperatures. We shall therefore leave unexplained the low temperature behaviour and concentrate on the behaviour in the range between 3.0° K and the transition temperature.

The problem raised by the smallness of the effect in this range may be understood most readily from a consideration of the entropy change accompanying the magnetization of a superconductor. That such an entropy change occurs is a thermodynamical consequence of the temperature variation of the penetration depth, through the relation

$$(\partial S/\partial H)_T = (\partial M/\partial T)_H,$$

where S is the entropy and M the magnetic moment of the specimen considered. If we consider a long specimen of zero demagnetizing coefficient, having volume V and surface area A , the magnetic moment in field H may be written in the form

$$M = -\frac{1}{4\pi}H(V - \lambda A).$$

Hence, neglecting thermal expansion,

$$(\partial S/\partial H)_T = \frac{1}{4\pi}AH(\partial\lambda/\partial T)_H. \quad (1)$$

The experimental results indicate that no great error will result from taking λ in this equation as independent of H , and we may therefore write λ as $\lambda_0(1 - t^4)^{-1}$, where $t = T/T_c$, (Daunt, Miller, Pippard & Shoenberg 1948) and integrate equation (1) to give the entropy change $\Delta S(H)$ when a field H is applied,

$$\Delta S(H) = \frac{vH^2}{4\pi T_c} \frac{t^3}{1 - t^4},$$

where v is written for λA , the volume of the penetration layer. The physical meaning of this entropy change is clearly that the stresses exerted by the magnetic field on the electron assembly result in a shift of the equilibrium configuration, and it is natural at first sight to assume that the equilibrium is shifted only in the surface layer into which the field penetrates. If then we take $\Delta S(H)$ as distributed in the layer of volume v the change in entropy density, $\Delta s(H)$, at the surface is given by the equation

$$\Delta s(H) = \frac{H^2}{4\pi T_c} \frac{t^3}{1 - t^4}.$$

In particular, at the critical field, H_c , which is given sufficiently precisely by $H_0(1 - t^2)$, the change in entropy density may be written

$$\Delta s_1 = \frac{H_0^2}{4\pi T_c} \frac{t^3(1 - t^2)}{1 + t^2}. \quad (2)$$

To understand the significance of this result, let us compare it with the difference in entropy density between the superconducting phase in zero field and the normal phase at the same temperature

$$\Delta s_2 = \frac{H_0^2}{2\pi T_c} t(1-t^2). \quad (3)$$

From equations (2) and (3)

$$\frac{\Delta s_1}{\Delta s_2} = \frac{1}{2} \frac{t^2}{1+t^2}.$$

Near the transition temperature, then, the change in entropy density in the surface layer amounts to one quarter of the entropy difference between the two phases, yet the corresponding change in penetration depth is only of the order of 1 %. It seems incredible that any theory should be able to account for this result, and it is much more likely that the experimental results should be interpreted as showing that there is no great change in entropy density on magnetization, in contradiction to equation (2). Since the change in total entropy is determined by thermodynamical arguments we are led to the conclusion that the entropy change is distributed, not in the penetration layer, but in a considerably thicker layer at the surface, so that the change in entropy density is correspondingly less. This hypothesis is consistent with the idea that the superconducting phase is one having the property of long-range order, as has been suggested on many occasions previously (for example, London 1948), in that it is not permissible to regard the constitution of the assembly as capable of variation over a short distance, but quite large elements of the assembly, of dimensions greater than λ , must rather be treated as units within which there is little or no variation.

In order to estimate the size of these elements, that is, the range of order, it is necessary to have some model which determines the relation between penetration depth and entropy. Although evidence is presented in B that the two-fluid model of Gorter & Casimir (1934) is not an entirely satisfactory representation of a superconductor, it does nevertheless correlate the thermodynamic properties with the temperature variation of penetration depth (Daunt *et al.* 1948), and there is thus hope that it will also be adequate for a discussion of the field variation. We shall apply this theory to a lamina of thickness a at the surface of a plane slab of superconductor having unit area, introducing the hypothesis of long-range order by means of the assumption that the constitution is uniform within the lamina. In the absence of a magnetic field the free energy is given on this theory by the equation

$$F_0 = -\frac{aH_0^2}{8\pi} \{\omega + 2t^2 \sqrt{(1-\omega)}\}, \quad (4)$$

where ω represents the fraction of superconducting electrons*: $\omega = 1$ when $t = 0$ and $\omega = 0$ when $t = 1$. Moreover ω is constant over the lamina and is related to the penetration depth by the equation

$$(\lambda_0/\lambda)^2 = \omega. \quad (5)$$

* ω is here used instead of Gorter & Casimir's $(1-x)$, and their parameter α is put equal to $\frac{1}{2}$, which is consistent with a parabolic critical field curve.

In a magnetic field, penetrating to a distance λ on one side of the lamina, the Gibbs function $\left(F - \frac{1}{4\pi} HB\right)$ may be written in the form,

$$G = -\frac{\alpha H_0^2}{8\pi} \{\omega + 2t^2 \sqrt{(1-\omega)}\} - \frac{\lambda H^2}{8\pi}.$$

If H and t are kept constant, ω takes such a value as to minimize G , and an equation for ω may be derived immediately,

$$1 - \sqrt{\left(\frac{1-\omega_0}{1-\omega}\right)} - \frac{\alpha_0 h^2}{2\omega^{\frac{3}{2}}} = 0, \quad (6)$$

where $\omega_0 = 1 - t^4$, the equilibrium value of ω in zero field, $h = H/H_0$, and $\alpha_0 = \lambda_0/a$. If $\alpha_0 \ll 1$, a sufficiently accurate solution of equation (6) is given by

$$1 - \frac{\omega}{\omega_0} = \frac{\alpha_0 h^2}{\frac{\omega_0^{\frac{3}{2}}}{1-\omega_0} - \frac{3}{2}\alpha_0 h^2},$$

and hence, by equation (5),

$$\frac{\Delta\lambda}{\lambda} = \frac{\alpha_0 h^2}{\frac{2\omega_0^{\frac{3}{2}}}{1-\omega_0} - 3\alpha_0 h^2}. \quad (7)$$

This formula may be applied to the results shown in figure 4 to deduce values of α_0 , and hence the depth over which ω may be taken as constant. For the reasons given above we shall consider only temperatures above 3.0° K, for which the following values of α_0 are deduced:

T (° K)	α_0
3.6	0.055
3.4	0.042
3.2	0.022

It would be unwise to interpret the variation of α_0 as evidence for variation with temperature of the range of order, on account both of uncertainties in the theoretical basis of the calculation and of doubts about the meaning of the behaviour below 3° K. However it is perhaps not unduly optimistic to take these results as indicating the existence of long-range order in superconducting tin over a range of about $20\lambda_0$, i.e. 10^{-4} cm. This distance is of course very much greater than the smallest specimen which may exhibit the superconducting properties of the bulk material; for example, Shoenberg's (1940) mercury colloids, though smaller in diameter than the penetration depth, nevertheless had almost the same transition temperature as bulk mercury. The range of order must therefore not be regarded as a minimum range necessary for the setting up of an ordered state, but rather as the range to which order will extend in the bulk material. It is perhaps helpful to imagine the electron assembly as analogous to a solid which can recrystallize very easily. Individual small crystallites will grow at the expense of their neighbours, and the process will only stop when large crystallites have formed or when impurities disturb the growth. So in the superconductor we may picture ordered regions growing in size until further growth is inhibited by impurities or thermal vibrations or, in small specimens, by the boundaries of the material.

(b) *Further evidence for long-range order*

The extreme sharpness of the second-order transition exhibited by pure superconductors in the absence of a magnetic field may be taken as further evidence that the superconducting state is one of long-range order, since if merely local order were possible one would expect it to persist to somewhat higher temperatures than T_c and to give rise to such a rounding of the specific heat anomaly as is observed to some extent in all other second-order transitions. The idea that the sharpness of the specific heat anomaly may be used to determine the volume of ordered regions was applied by Keesom & Keesom (1935) to the lambda transition in liquid helium. Essentially their argument involves the division of the liquid into domains, each of constant temperature but with a variation of temperature from one domain to another as a result of fluctuations. The specific heat of each domain is assumed to be given by an ideal curve having a discontinuity at the critical temperature and the resulting specific heat of the whole assembly is then easily calculated. It is difficult, however, to justify this mixture of statistical methods which assumes that, in spite of fluctuations in the assembly as a whole, the entropy of a single domain is that appropriate to the most probable rather than the average configuration. Possibly no great error arises from this assumption, but in the absence of a detailed model the error cannot be estimated.

The same difficulty would arise if we attempted to apply the argument to the specific heat curve of a superconductor, but it is not so acute, or at any rate the nature of the assumption is more easily realized, if we consider instead the resistance curve. The difficulty with the specific heat arises from the fact that a domain at a given temperature may take up a large variety of different complexions, each with its appropriate entropy, and that without a detailed knowledge of the complexions we cannot estimate the mean entropy. For the resistance, however, a different situation obtains. The occurrence of the Meissner effect shows that superconductivity is not to be regarded as the limiting case of good conductivity; moreover, in the normal state the resistance is sensibly independent of temperature. These two facts suggest that the complexions may be divided into two groups, those which confer the property of superconductivity on the domain and those which involve normal resistive properties. The entropy may show a continuous gradation between the two extremes but the resistance has no intermediate value. Now, a domain having an energy corresponding to the mean energy density at the transition temperature has an equal probability of being in a normal or superconducting state, and we shall assume that if its energy is less than this it will be superconducting, and if greater it will be normal. Let us now for simplicity imagine the superconductor divided into equal domains, each of volume v , and determine the shape of the resistance curve.

If the mean temperature is T , the fraction of domains having energies between E and $E + dE$ is given by the expression,

$$f(E) dE = \frac{\beta}{\sqrt{\pi}} \exp[-\beta^2(E - \bar{E})^2] dE,$$

where \bar{E} is the mean energy at temperature T , $\beta^2 = 1/CvkT^2$, and C is the specific heat per unit volume. If $T = T_c + \tau$, the fraction of domains in the superconducting

state is, according to the above assumption, the fraction $P(\tau)$ whose energy is less than the mean energy by an amount greater than $C_n v \tau$, C_n being the specific heat per unit volume of normal material, i.e.

$$P(\tau) = \frac{\beta C_n v}{\sqrt{\pi}} \int_{\tau}^{\infty} \exp[-\beta^2 C_n^2 v^2 \tau^2] d\tau = \frac{1}{2}[1 - \operatorname{erf}(\gamma\tau)], \quad (8)$$

where

$$\gamma = \beta C_n v = \sqrt{\frac{C_n v}{2kT^2}}$$

Now the mean conductivity of a material of which a fraction P is in the form of approximately spherical superconducting inclusions, the rest having conductivity σ_n , may be calculated by a method exactly analogous to Lorentz's calculation of the internal field correction in polarizable media, yielding the result

$$\bar{\sigma} = \sigma_n \frac{1 + 2P}{1 - P}.$$

Although this expression tends to infinity as P tends to unity it cannot be expected to be more than a rough approximation for the larger values of P . Substituting equation (8) we arrive at an expression for the temperature variation of $\bar{\sigma}$,

$$\bar{\sigma} = 2\sigma_n \frac{2 - \operatorname{erf}(\gamma\tau)}{1 + \operatorname{erf}(\gamma\tau)}. \quad (9)$$

This is a typical sigmoid curve whose width is determined by γ .

The sharpest resistance transition recorded was found by de Haas & Voogd (1931) for a single crystal of tin, for which the conductivity rose from $1.14\sigma_n$ to $1.73\sigma_n$ in a temperature range of only 10^{-3}°K . Using this result in equation (9) we derive a value of 600 deg.^{-1} for γ , and hence a value for v of $9.5 \times 10^{-14} \text{ cm.}^3$. A sphere of this volume would have a diameter of $5.7 \times 10^5 \text{ cm.}$, i.e. about $10\lambda_0$.

The sharpness of the second-order transition thus leads to an estimate of the domain size of the same order of magnitude as, but slightly smaller than, that obtained from the field variation of the penetration depth. It should be realized that this estimate is, for two reasons, likely to be too small. First, if the assumption that domains with energy less than that corresponding to the transition temperature are always superconducting is seriously in error, an improbable contingency, it will mean that the spread of the resistance transition for a given domain size has been underestimated. Secondly, the observed transition may be broadened as a result of effects other than fluctuations, such as strains. On the other hand, the fact that a measuring current is needed in order to determine the transition may result in a narrowing of the transition curve, since the magnetic field of the current will change the second-order into a first-order transition and reduce the probability of fluctuations. It is unlikely, however, that this was a serious effect in the experiment quoted, since the field of the measuring current was small enough not to cause a shift of the transition temperature by more than about 10^{-3}°K . We may therefore take this result as confirmation of the view that the superconducting phase of pure tin has an ordered structure over a range of approximately 10^{-4} cm.

I should like to express my gratitude to Dr H. London for drawing my attention, several years ago, to this problem, to Professor R. Peierls and Dr D. Shoenberg for their critical interest, and to Mr R. G. Chambers for help with the experiments.

REFERENCES

- Andrew, E. R. 1948 *Proc. Roy. Soc. A*, **194**, 80.
 Daunt, J. G., Miller, A. R., Pippard, A. B. & Shoenberg, D. 1948 *Phys. Rev.* **74**, 842.
 Désirant, M. & Shoenberg, D. 1948 *Proc. Roy. Soc. A*, **194**, 63.
 Ginsburg, V. 1947 *J. Phys. U.S.S.R.* **11**, 93.
 Gorter, C. J. & Casimir, H. B. G. 1934 *Z. Tech. Phys.* **12**, 539.
 de Haas, W. J. & Voogd, J. 1931 *Commun. Phys. Lab. Univ. Leiden*, no 214c.
 Heisenberg, W. 1947 *Z. Naturforsch.* **2a**, 185.
 Keesom, W. H. & Keesom, A. P. 1935 *Physica*, **2**, 557.
 Koppe, H. 1949 *Z. Naturforsch.* **4a**, 74.
 Landau, L. 1941 *J. Phys. U.S.S.R.* **5**, 71.
 Lasarev, B. G., Galkin, A. A. & Khotkevich, V. I. 1947 *C.R. Acad. Sci. U.R.S.S.* **55**, 805.
 von Laue, M. 1949 *Ann. Phys., Lpz.*, **5**, 197.
 Laurmann, E. & Shoenberg, D. 1949 *Proc. Roy. Soc. A*, **198**, 560.
 London, F. 1948 *Phys. Rev.* **74**, 562.
 London, H. 1936 *Proc. Roy. Soc. A*, **155**, 102.
 Macdonald, D. K. C. 1949 *Nature*, **163**, 639.
 Pippard, A. B. 1950a *Proc. Roy. Soc. A*, **203**, 98.
 Pippard, A. B. 1950b *Proc. Roy. Soc. A*, **203**, 195.
 Shoenberg, D. 1940 *Proc. Roy. Soc. A*, **175**, 49.

Size effect variation of the electrical conductivity of metals

BY D. K. C. MACDONALD, *Clarendon Laboratory, Oxford*, AND
 K. SARGINSON, *Somerville College, Oxford*

(Communicated by F. E. Simon, F.R.S.—Received 28 March 1950)

This paper records experiments and theoretical work concerned with the variation of conductivity with size in metals.

Experimental results for the conductivity in thin wires of pure sodium of varying diameter in the absence of a magnetic field and also in the presence of longitudinal and transverse magnetic fields are given.

Using the general statistical theory of metals the variation of resistance with size in the case of conductivity wires of square cross-section is calculated for comparison with the first set of experiments. A theoretical investigation follows of the alteration in conductivity produced in metallic films by the application of transverse magnetic fields, and this is compared with the corresponding experimental results obtained on the sodium cylinders.

1. INTRODUCTION

Towards the end of the last century, it was known that very thin films of metal exhibited a higher specific electrical resistance than the same metal in bulk. In particular, Miss I. Stone (1898) carried out a detailed experimental study of films of silver deposited by the 'Rochelle salt process'. J. J. Thomson (1901) was the first to suggest that the source of the phenomenon lay in the limitation of the mean free

path of the conduction electrons by the geometry of the specimen when its size became sufficiently small to be of the order of the mean free path due to electron-lattice collisions. However, when we remark that the mean free path of electrons in, say, bulk silver is $\sim 10^{-5}$ cm. at 0°C , it becomes obvious that very thin films of metal will be required for experiments at ambient temperatures. This limitation in turn calls forth a number of subsidiary resistive effects, often time dependent, which arise from the mode of preparation of such thin films. These effects, although themselves of interest (see Andrew 1949 for references), naturally tend to obscure the pure geometrical limitation of free path which it is desired to examine. On the other hand, the advent of low temperatures has enabled fresh work to be undertaken on relatively large specimens, since the conductivity of pure metals increases greatly at low temperatures. Lovell (1936) examined films of rubidium at temperatures around 70°K , attained by pumping liquid oxygen and nitrogen. These films were still rather thin ($\sim 40\text{ \AA}$), since the increase of mean free path in bulk at such a temperature is only about four times the room-temperature values. However, Andrew (1949) has recently extended the field to the liquid-hydrogen and liquid-helium region (20°K and below). Pure tin, as used by Andrew, may drop in electrical resistance at temperatures $\sim 2^\circ\text{K}$ to lower than 10^{-4} of its room temperature value, indicating a bulk electronic mean free path $\sim 10^{-2}$ cm. Consequently, appreciable size-effect variation may now be expected to occur in specimens as thick as 100μ (10^{-2} cm.). Andrew examined pure mercury in capillaries down to $\sim 6\mu$ diameter and tin in foils as thin as $\sim 3\mu$ and found generally satisfactory agreement with theory. Part of the latter he developed himself to give an approximate solution for the size-effect in cylinders, previous theoretical work (Thomson 1901, Lovell 1936; Fuchs 1938) having been applied to planar models ('films').

All the theoretical investigations have assumed the simple isotropic free-electron model of a metal. Since neither tin nor mercury specifically are simple metals, it was felt worth while to undertake work on sodium which as an alkali metal might be expected *a priori* to approximate to the ideal metal and has also been found (MacDonald & Mendelssohn 1948, 1950; MacDonald 1950) through detailed low-temperature work to obey almost perfectly the free-electron model. Size-effect variation has been observed at temperatures between 20 and 4.2°K on very pure sodium in capillaries between 15 and 66μ diameter, and theory has been developed, extending Fuchs's analysis, for comparison with these experiments. A new phenomenon has also been observed; the application of transverse and longitudinal magnetic fields produces a marked variation of electrical resistance under these conditions. In particular, a longitudinal magnetic field produces a monotonic decrease of the resistance which tends towards the bulk value for large magnetic fields. These effects would be expected to arise if one considers the alteration produced in the shape of the electron orbits in relation to the size of the specimen.

The discovery of the effect has occasioned theoretical investigation by Sondheimer (1949) and Chambers (1950).^{*} We ourselves have considered the case of a transverse magnetic field applied in the plane of a film, since this situation appeared to us closest to the experimental conditions involved. A preliminary note appeared

^{*} We are grateful to Mr Chambers for letting us see his manuscript before publication.

in *Nature* (Sarginson & MacDonald 1949). Good qualitative agreement is found with the experiments, but it appears still desirable to obtain experimental data in the future on sodium *films* in magnetic fields in order to test the theory more closely.

The investigation also provides some information on the variation of the Hall field across the film.

2. EXPERIMENTS. BY D. K. C. MACDONALD

2.1. Preparation of specimens

Very pure sodium, obtained through the courtesy of Messrs British Thomson-Houston Co. Ltd, was used throughout. Experiments on 'large' specimens (MacDonald & Mendelssohn 1948, 1950) showed this to have a residual resistance

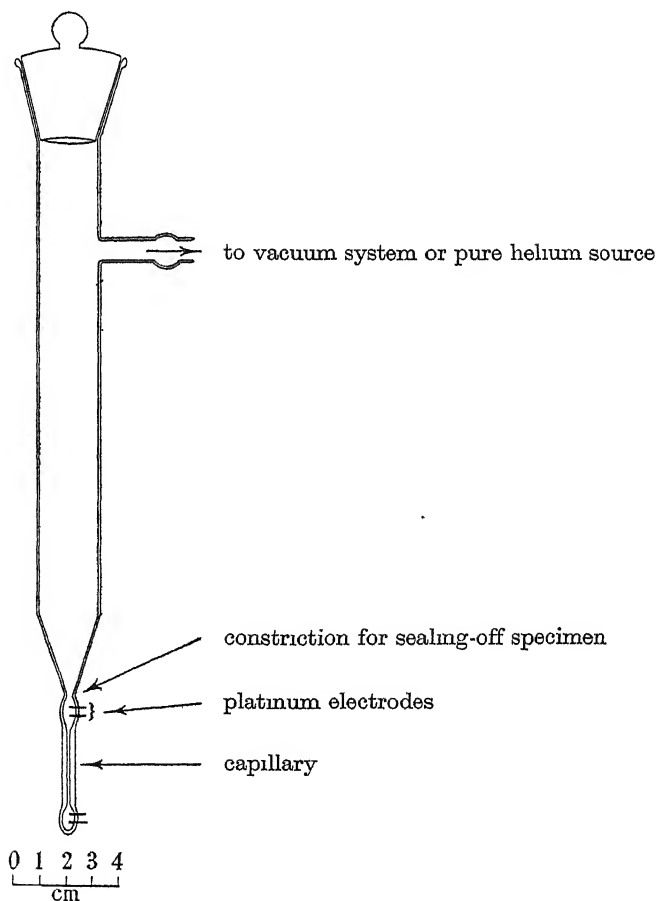


FIGURE 1. Typical capillary mould.

$\sim 7.2 \times 10^{-4}$ of its room-temperature value when measured in liquid helium (4.2°K). The specimens were prepared by melting sodium *in vacuo* in a soft-glass system (see figure 1) and then pressing them into soft-glass capillary moulds by pure helium gas at 1 atm. pressure. The capillary specimens were provided with platinum potential and current electrodes at either end. After solidification the specimens were cracked off at the constriction and then immediately sealed with a suitable compound.

It has been found that capillaries down to $\sim 30\mu$ diameter can be filled quite readily with sodium, but very considerable difficulty was experienced with the narrower capillaries. Since sodium 'wets' clean glass this difficulty is not to be ascribed to surface-tension effects as was the case in Andrew's experiments on mercury. This was in fact verified, since pressurizing with helium at ~ 4 atm. did not aid in trying to fill 20μ capillaries. It was ultimately found that, owing to the high chemical activity of the sodium, the utmost cleanliness and freedom from any form of contamination were essential to permit filling. Thus, when a specimen was allowed

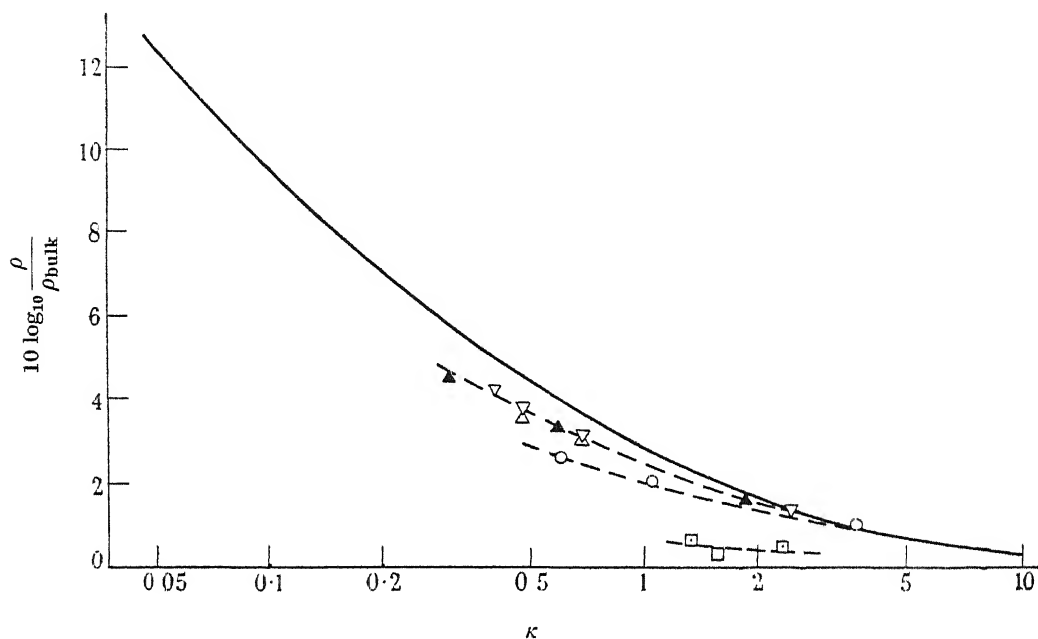


FIGURE 2. Size effect of electrical resistance in sodium wires of varying diameter: \blacktriangle , 15μ ; \triangle , ∇ , 20μ ; \circ , 30μ ; \square , 66μ capillaries. —, theoretical curve.

to lie for an hour after glass-blowing (the open end being plugged with cotton-wool), moisture from the atmosphere condensed into the capillary creating a microscopic deposit sufficient to prevent the subsequent passage of liquid sodium, although the mould was 'baked out' before filling under high vacuum at temperatures $\sim 150^\circ\text{C}$. Success was achieved through a specific glass-blowing technique* designed to prevent any condensation in the capillary in the process and by attaching the apparatus immediately it was complete to the pump system.

Some experiments were also made on tin, gold and silver foils and these were rolled down to thicknesses $\sim 6\mu$ from spectrographically pure rod or wire obtained from Messrs Johnson, Matthey and Co. Ltd. The rollers were carefully cleaned before use to avoid any surface contamination.

2.2. Experimental methods

Measurements were carried out in both a Simon expansion helium liquefier (Simon 1936) and also a simple hydrogen cryostat for some measurements between ~ 11 and

* I am indebted to Mr D. G. Saxton for his patience and trouble in this aspect of the work

20° K. Temperatures in both cases were measured with a helium gas thermometer. The electrical resistance was measured in some cases with a Tinsley potentiometer and in the wider specimens with a high-impedance galvanometer amplifier (MacDonald, 1947) using negative feed-back.

2.3. *Experimental results*

The experimental observations on the simple size effect are shown in figure 2. The results on the various specimens are plotted as the broken lines with the experimental points; the theoretical curve (to be discussed below) is shown as the full line.* Figures 5 and 6 show the results obtained for the change of conductivity under transverse and longitudinal magnetic fields which are to be compared with the theory.

3. THEORY. BY K. SARGINSON AND D. K. C. MACDONALD

3.1. *Simple size-effect in a square wire*

3.1.1. *General*

Thomson (1901) and Lovell (1936) both analyzed the increase of resistance in a thin film using the concept of limitation of mean free path by the dimensions of the conductor; the method is fundamentally similar to that used in the elementary kinetic theory analysis of transport phenomena. Nordheim (1931, 1934), in considering the case of a thin wire, essentially adapted an expression appropriate to the flow of gas down a tube of diameter comparable to the molecular mean free path. Andrew (1949) treated the case of a thin wire approximately by an extension of the Thomson-Lovell method.

Fuchs (1938), however, approached the problem of a thin film using the full statistical theory appropriate to an electron gas in dynamical equilibrium. This method dates principally from the time of Lorentz and has become a familiar feature of the modern theory of metals. The derivation of a (probability) distribution function of velocity and position yielding statistical equilibrium under the influence of the applied (electric ' E ' and magnetic ' H ') fields and the electron-lattice collisions is basic to this treatment. The expression of this requirement leads in general to an integro-differential equation for the distribution function, N . This is known as the Boltzmann equation. Given the distribution function as a solution, the electric current (and hence the electrical conductivity) and the heat current (if any) may be derived, essentially as velocity moments. In the case of limitation by size, the scattering at the physical edges enters as appropriate boundary conditions to be imposed on the solution.

More recently, Dingle (1950)† has been dealing in this way with a *cylindrical* wire, while Chambers (1950) has used generalized kinetic theory methods to treat the same problem. We, however, have extended Fuchs's treatment to the analysis of a *square* wire of side a . In all these treatments the electron-lattice collisions are assumed to be

* The abscissa is the ratio of the specimen diameter as determined by the bore of the capillary to the electron mean free path in bulk metal. The latter is known from detailed low temperature experiments on a 'large' specimen of the same material.

† We have not yet had the opportunity of reading Dr Dingle's paper.

sufficiently represented by a simple isotropic relaxation time, τ , throughout. This corresponds precisely to the assumption mentioned earlier of the idealized quasi-free electron model of a metal. The introduction of this relaxation time, related to the bulk conductivity, σ by the equation $\sigma = \frac{\nu e^2 \tau}{m}$ (cf. Wilson 1936) (where ν is the number of free electrons per unit volume), reduces the Boltzmann equation to a partial differential equation. It is also generally assumed, following Lorentz, that the distribution function, N , under the applied fields differs only slightly from that, N_0 , existing under true equilibrium conditions ($E = H \equiv 0$). This corresponds to the fact that, in metals, the average energy gained by an electron from the fields between collisions with the lattice is very small compared with the intrinsic Fermi energy. Finally, it is assumed that no restriction need be made on the magnitude of H in the solution of the Boltzmann equation. This corresponds to neglect of possible quantization of the electron orbits in the magnetic field (cf. Țițeica 1935; Wilson 1936), and is therefore in harmony with a purely classical treatment of the metal. Since sodium in bulk exhibits a very small magneto-resistive effect it would appear that the neglect of quantization is then justifiable, but caution must be exercised when considering more complex metals.

3.12. Detailed analysis

Let the sides of the square cross-section of the wire be a . Choose axes Ox , Oy , Oz with Ox , Oy along sides of the square and Oz along one edge of the wire.

The appropriate Boltzmann equation is

$$v_x \frac{\partial n}{\partial x} + v_y \frac{\partial n}{\partial y} + \frac{n}{\tau} = \frac{eE}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v}, \quad (1)$$

where

$$n = N - N_0,$$

$N(x, y, z, v_x, v_y, v_z)$ is the actual distribution function,

$N_0(v_x, v_y, v_z)$ being that with $E = 0$,

and

v is the velocity of the electrons on the Fermi surface.

The general solution of (1) is

$$n = \frac{eE}{m} \frac{\tau v_z}{v} \frac{\partial N_0}{\partial v} \left(1 - e^{-x/\tau v_x} f\left(\frac{y}{v_y} - \frac{x}{v_x}\right) \right), \quad (2a)$$

or equally well

$$n = \frac{eE}{m} \frac{\tau v_z}{v} \frac{\partial N_0}{\partial v} \left(1 - e^{-y/\tau v_y} \phi\left(\frac{x}{v_x} - \frac{y}{v_y}\right) \right), \quad (2b)$$

where f and ϕ are arbitrary functions to be determined from the boundary conditions. If we assume inelastic electron scattering at the boundaries, then we require that n be independent of the ratio $v_x : v_y : v_z$ at $x = 0, a$; $y = 0, a$ for those electrons *leaving* the boundaries. This can only be realized if $n = 0$ for these conditions. The complete distribution function can then be most simply represented by the scheme of table 1 in velocity space.

TABLE 1

$n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-y/\tau v_y}),$ $y < -\frac{v_y}{v_x} (a - x)$ $n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-(a-x)/\tau v_x}),$ $y > -\frac{v_y}{v_x} (a - x)$	$n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-y'/\tau v_y}),$ $y < \frac{v_y}{v_x} x.$ $n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-x'/\tau v_x}),$ $y > \frac{v_y}{v_x} x$
$n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-(a-x)/\tau v_x}),$ $(a - x) < \frac{v_x}{v_y} (a - y)$ $n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-(a-y)/\tau v_y}),$ $(a - x) > \frac{v_x}{v_y} (a - y)$	$n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-x'/\tau v_x}),$ $x < -\frac{v_x}{v_y} (a - y)$ $n = \frac{eE\tau}{m} \frac{v_z}{v} \frac{\partial N_0}{\partial v} (1 - e^{-(a-y)/\tau v_y}),$ $x > -\frac{v_x}{v_y} (a - y)$

It is obvious that each quadrant in velocity space will contribute equally to the total current J_z ; thus

$$\begin{aligned}
 J_z &= \frac{4}{a^2} \iiint_{\substack{v_x > 0 \\ v_y > 1}} \left(\frac{e^2 E \tau}{m v} \frac{\partial N_0}{\partial v} v_z^2 \int_{x=0}^a \left\{ \int_0^{(v_y/v_x)x} (1 - e^{-y'/\tau v_y}) dy' \right. \right. \\
 &\quad \left. \left. + \int_{(v_y/v_x)x}^a (1 - e^{-x'/\tau v_x}) dx' \right\} dv_x dv_y dv_z \right. \\
 &\quad \left. + \frac{4}{a^2} \iiint_{\substack{v_x > 0 \\ v_y/v_x > 1}} \left(\frac{e^2 E \tau}{m v} \frac{\partial N_0}{\partial v} v_z^2 \int_{y=0}^a \left\{ \int_0^{(v_x/v_y)y} (1 - e^{-x'/\tau v_x}) dx' \right. \right. \right. \\
 &\quad \left. \left. + \int_{(v_x/v_y)y}^a (1 - e^{-y'/\tau v_y}) dy' \right\} dv_x dv_y dv_z \right) \quad (3) \\
 &= \frac{4}{a^2} \left(a^2 \iiint_{\substack{v_x > 0 \\ v_y > 0}} \frac{e^2 E \tau}{m v} \frac{\partial N_0}{\partial v} v_z^2 dv_x dv_y dv_z - 2 \iiint_{0 < v_x < v_y} \frac{e^2 E \tau}{m v} \frac{\partial N_0}{\partial v} v_z^2 \left\{ a \tau v_y (1 - e^{-a/\tau v_y}) \right. \right. \\
 &\quad \left. \left. + a \tau v_x (1 + e^{-a/\tau v_y}) - 2 \tau^2 v_x v_y (1 - e^{-a/\tau v_y}) \right\} dv_x dv_y dv_z \right). \quad (4)
 \end{aligned}$$

Thus

$$\begin{aligned}
 \frac{\sigma}{\sigma_{\text{bulk}}} &\equiv \frac{\rho_{\text{bulk}}}{\rho} \\
 &= 1 - \frac{2 \iiint_{0 < v_x < v_y} v_z^2 \left(\frac{\tau v_y}{a} \{1 - e^{-a/\tau v_y}\} + \frac{\tau v_x}{a} \{1 + e^{-a/\tau v_y}\} - \frac{2 \tau^2 v_x v_y}{a^2} \{1 - e^{-a/\tau v_y}\} \right) dv_x dv_y dv_z}{\iiint_{\substack{v_x > 0 \\ v_y > 0}} v_z^2 dv_x dv_y dv_z}. \quad (5)
 \end{aligned}$$

The method to be adopted in evaluating (5) depends on the value of the parameter $\kappa \equiv a/l$, where l , the mean free path, $= v\tau$, by definition.

For κ large (i.e. incipient size effect), we may express (5) as

$$\frac{\sigma}{\sigma_{\text{bulk}}} = 1 - \frac{6}{\pi} \left(\frac{\pi}{8\kappa} - \frac{2}{15\kappa^2} \right) + \frac{12}{\pi\kappa} \int_{\theta=0}^{\frac{1}{2}\pi} \int_{\phi=0}^{\frac{1}{2}\pi} e^{-\kappa(\sin\theta \cos\phi)} \cos\phi \sin^2\theta \cos^2\theta d\theta d\phi \\ + \frac{12}{\pi\kappa^2} \int_0^{\frac{1}{2}\pi} \left(\frac{1}{2} e^{-\sqrt{2}\kappa(\sin\theta)} - e^{-\kappa(\sin\theta)} \right) \sin^3\theta \cos^2\theta d\theta. \quad (6)$$

By then considering upper bounds for the latter integral terms we have for $\kappa \geq 4$, with an upper limit of error $\sim 2/1000$,

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx 1 - 0.75/\kappa + 0.254/\kappa^2. \quad (7)$$

It is interesting to compare (7) with the expression obtained by Fuchs for a *film* of thickness a under the same condition, namely,

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx 1 - 0.375/\kappa + \dots \quad (7a)$$

Comparison of the terms in $1/\kappa$ in (7) and (7a) shows the independently additive limitation imposed by the two boundaries in the square wire in the limit of incipient size effect.

On the other hand, for small κ (intense size effect) proceeding from (5) and (6) we may obtain an expansion in powers of κ , which finally yields the approximate formula for $\kappa \ll 1$:

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx \frac{3}{4}\kappa \log_e \left(\frac{\sqrt{2}+1}{\sqrt{2}-1} \right) - \frac{1}{2}\kappa (\sqrt{2}-1) \\ \approx 1.115\kappa. \quad (8)$$

It is again interesting to compare Fuchs's corresponding expression for a thin *film*:

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx \frac{3\kappa}{4} \log_e \left(\frac{1}{\kappa} \right). \quad (8a)$$

The much more rapid decrease of conductivity in the case of the 'wire' is immediately evident.

We further note that Andrew's (1949) formula for a cylindrical wire of diameter b under the condition $\kappa' \equiv b/l, \ll 1$, is simply

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx \kappa'. \quad (8b)$$

If now we set $b = \frac{2}{\sqrt{\pi}}a$ so that the cross-sectional areas in Andrew's and our cases are equal we have from (8b)

$$\frac{\sigma}{\sigma_{\text{bulk}}} \approx 1.128\kappa, \quad (8c)$$

in very close agreement with (8).

Finally, for values of $\kappa \sim 1$, we may continue from (5) to obtain single integrals tolerably convenient for direct numerical integration. We find

$$\begin{aligned} \frac{\sigma}{\sigma_{\text{bulk}}} = 1 - \frac{3}{\pi} \left(\frac{\pi}{4\kappa} - \frac{4}{15\kappa^2} \right) + \frac{12}{\pi} \int_0^{\frac{1}{2}\pi} \left(\frac{\pi}{4\kappa} \cos \theta - \frac{1}{3\kappa} \sin \theta - \frac{1}{3\kappa^2} \sin 2\theta \right) e^{-\kappa/(\cos \theta)} \sin^3 \theta d\theta \\ + \frac{12}{\pi} \int_{\frac{1}{2}\pi}^{\pi} \left\{ \frac{\cos \theta}{\kappa} \left(\frac{1}{4}\pi - \frac{1}{2} \cos^{-1}(\cot \theta) \right) + \frac{\cot^2 \theta}{2\kappa} \sqrt{(-\cos 2\theta)} - \frac{1}{3\kappa} \left(\sin \theta - \frac{(-\cos 2\theta)^{\frac{1}{2}}}{\sin^2 \theta} \right) \right. \\ \left. - \frac{2}{3\kappa^2} \left(\sin \theta \cos \theta - \frac{\cos \theta}{\sin^2 \theta} (-\cos 2\theta)^{\frac{1}{2}} \right) \right\} e^{-\kappa/(\cos \theta)} \sin^3 \theta d\theta. \quad (9) \end{aligned}$$

For $\kappa = 1$ we find $\sigma/\sigma_{\text{bulk}} = 0.517$, and for $\kappa = 0.33$ we find $\sigma/\sigma_{\text{bulk}} = 0.277$.

3.13. Comparison with experiment

In figure 2, using (7), (8) and (9), we have plotted $10 \log_{10} \rho/\rho_{\text{bulk}}$ against κ . The experimental data obtained at various temperatures between 4.2 and 20.4° K on cylindrical specimens of diameter 15, 20 (two specimens), 30 and 66 μ are also plotted thereon; the diameters, b , have been converted to an effective a using the relation $a = \frac{\sqrt{\pi}}{2} b$ as above.

We notice immediately that all the experimental points lie below, or on, the theoretical curve, suggesting strongly that no impurity additional to that in a bulk specimen is present which would yield an artificially high ('residual') resistance at low temperatures. The general feature that all the results tend to lie below the theoretical curve suggests that the hypothesis of entirely diffuse scattering at the boundaries is not generally fulfilled. Furthermore, the uniform trend towards closer agreement with theory with reduction of diameter suggests strongly that the degree of elastic scattering is a function of the specimen size.

Two alternatives present themselves. On the one hand, it is possible that this is a property inherent in the size of the specimen itself, such that the uniformity of the crystalline structure of the surface depends in some way significantly on the volume of the sample. In order to examine such a hypothesis it would first be necessary to determine the surface conditions required to provide effective elastic scattering in a particular case. In the first place, if the surface were presumed irregular within a range less than the 'extent' of the electron then it seems evident that wholly diffuse scattering must ensue. If we turn to the uncertainty relation $\Delta x \Delta p_x \sim \hbar$ to determine Δx , we might set $\Delta p \sim p$, where p is the magnitude of the momentum on the Fermi surface. In this case $\Delta x \sim \lambda$, the de Broglie wave-length, and $\lambda \approx 7 \times 10^{-8}$ cm. for sodium, assuming one free electron per atom. Alternatively, one might set $\Delta p \sim \sqrt{(2mkT)}$, for $T \sim 10^\circ$ K this yields $\Delta x \approx 4 \times 10^{-6}$ cm. Both of these values are, however, very small compared with the size of our specimens, and therefore a *correlation* in this sense seems very improbable.

On the other hand, since the specimens were formed in drawn glass capillaries it appears possible that the intrinsic surface irregularities of the glass might be correlated with the diameter to which the capillary is drawn and hence impressed on the metal surface.

3.2. *Influence of a transverse magnetic field on the electrical conductivity of a film*3.21. *The distribution function*

In this section we investigate the change in conductivity produced in a thin film under the action of a transverse magnetic field in the plane of the film. Let a right-handed set of axes be taken so that the x -axis is in the direction of the applied electric field E , the y -axis is in the direction of the magnetic field H , and the z -axis is perpendicular to the plane of the film. The Hall field will be in the direction of the z -axis. It will not be constant as in cases hitherto considered (e.g. Sondheimer 1949, Chambers 1950), but will vary in magnitude across the film. Let, then, the Hall field be $F(z) + \frac{El}{r}$, where $\frac{El}{r}$ is the Hall field in bulk metal, r being the radius of the classical orbit of a free electron in the magnetic field. Further, in velocity space, we introduce polar angles χ, ψ such that $v_x = v \sin \chi \sin \psi$, $v_y = v \cos \chi$, $v_z = v \sin \chi \cos \psi$. The Boltzmann equation now reads

$$v \sin \chi \cos \psi \frac{\partial N}{\partial z} + \frac{eH}{mc} \frac{\partial N}{\partial \psi} - \frac{1}{\tau} \left(-N + \frac{1}{4\pi} \iint N(v, z) d\omega \right) = \frac{e}{m} \frac{\partial N}{\partial v} \left(E \sin \chi \sin \psi + \left\{ F(z) + \frac{El}{r} \right\} \sin \chi \cos \psi \right), \quad (10)$$

where the integral involved is to be taken over the Fermi surface, $d\omega$ being an element of solid angle.

The term $\frac{1}{\tau} \left(-N + \frac{1}{4\pi} \iint N(v, z) d\omega \right)$ gives the net change in N per unit time due to collisions. τ is the probability per unit time of an electron being scattered, and we assume that this has the same value as in bulk metal. In the case of bulk metal the integral term vanishes on the grounds of symmetry and τ may then be identified with the relaxation time (cf. Wilson 1936, p. 159).]

We assume, as usual, that N can be expressed as $N = N_0 + n$, where n is small. Then

$$v \sin \chi \cos \psi \frac{\partial n}{\partial z} + \frac{eH}{mc} \frac{\partial n}{\partial \psi} + \frac{n}{\tau} = \frac{e}{m} \frac{\partial N_0}{\partial v} \left(E \sin \chi \sin \psi + F(z) \sin \chi \cos \psi + \frac{El}{r} \sin \chi \cos \psi \right) + \frac{1}{4\pi} \iint n(v, z) d\omega.$$

Let $n = \frac{1}{v} \frac{e}{m} \frac{\partial N_0}{\partial v} \phi(v_x, v_y, v_z, z)$, and $\frac{mc v}{eH} = r$, the orbit radius, giving

$$\sin \chi \cos \psi \frac{\partial \phi}{\partial z} + \frac{1}{r} \frac{\partial \phi}{\partial \psi} + \frac{\phi}{l} = E \sin \chi \sin \psi + F(z) \sin \chi \cos \psi + \frac{El}{r} \sin \chi \cos \psi + \frac{1}{4\pi l} \iint \phi(v, z) d\omega. \quad (11)$$

Let the excess electron charge density at the point z be $\rho(z)$, then

$$\begin{aligned} \rho(z) &= -e \iiint n dv_x dv_y dv_z \\ &= \frac{2e^2 m^2 v}{h^3} \iint \phi(v, z) d\omega, \end{aligned} \quad (12a)$$

and Poisson's equation requires

$$\frac{\partial F(z)}{\partial z} = 4\pi\rho = \frac{8\pi e^2 m^2 v}{h^3} \iint \phi(v, z) d\omega. \quad (12b)$$

Thus ϕ must satisfy

$$\sin \chi \cos \psi \frac{\partial \phi}{\partial z} + \frac{1}{r} \frac{\partial \phi}{\partial \psi} + \frac{\phi}{l} = E \sin \chi \sin \psi + \frac{El}{r} \sin \chi \cos \psi + F(z) \sin \chi \cos \psi + \gamma l \frac{\partial F}{\partial z},$$

where $\gamma = \frac{h^3}{32\pi^2 e^2 l^2 m^2 v}$, $\sim 10^{-11}$, when $l \approx 45 \times 10^{-4}$ cm., $v \approx 10^8$ cm./sec. (the appropriate values for our sodium metal at $\sim 4^\circ$ K). Letting $z/r = \sin \chi \sin \eta - \xi$ and $\psi = \eta$, we have

$$\begin{aligned} \frac{1}{r} \frac{\partial \phi}{\partial \eta} + \frac{\phi}{l} &= E \sin \chi \sin \eta \\ &+ \frac{El}{r} \sin \chi \cos \eta + F(r \sin \chi \sin \eta - r\xi) \sin \chi \cos \eta + \gamma l F'(r \sin \chi \sin \eta - r\xi), \end{aligned}$$

which has as a general solution

$$\begin{aligned} \phi &= El \sin \chi \sin \eta + r e^{-(r/l)\eta} \int_{\eta}^{\eta} \{F(r \sin \chi \sin t - r\xi) \sin \chi \cos t \\ &\quad + \gamma l F'(r \sin \chi \sin t - r\xi)\} e^{(r/l)t} dt, \end{aligned} \quad (13)$$

where the lower limit of integration is arbitrary.

The solutions which satisfy the assumed boundary conditions, namely diffuse scattering at the edges, are set out in table 2 according to their regions of validity in co-ordinate and velocity space. The value of the inverse sine in the expressions is to be taken in the range $-\pi/2$ to $+\pi/2$.

In the solutions, A is a constant to be determined by the condition that $F(z)$ must vanish at $z = 0, d$. From table 2, when $z = 0$, $\phi = A$ for $v_z > 0$. From (12a), the excess density of electrons leaving $z = 0$ is therefore $\frac{2e^2 m^2 v}{h^3} 2\pi A$, and the total excess density of electrons at $z = 0$ is twice this. Thus

$$\left(\frac{\partial F(z)}{\partial z}\right)_{z=0} = \frac{32\pi^2 e^2 m^2 v}{h^3} A = \frac{A}{\gamma l^2};$$

we note that A is an odd function of H .

The function Φ_2 gives the distribution at points in the film for values of χ and ψ such that electrons scattered from either edge could not directly reach the point with such values of χ and ψ . The range concerned is determined by the geometry of the film coupled with the appropriate classical orbits. The arbitrary limit of integration in this function is determined by the condition that Φ_2 must assume the same value at $\psi = -\frac{1}{2}\pi$ and $\psi = \frac{3}{2}\pi$ for a given χ and z .

3.22. The Hall field

In order to obtain a complete solution for the conductivity it is necessary to determine the Hall field as a function of z and H universally. In principle this

TABLE 2. SOLUTION OF THE TRANSPORT EQUATION

I: $z < r \sin \chi (1 + \sin \psi)$, $d - z > r \sin \chi (1 - \sin \psi)$

$$\phi = El \sin \chi \sin \psi + \Phi_1,$$

where

$$\begin{aligned} \Phi_1 = r e^{-(r/l)\psi} \int_{\sin^{-1}\left(\sin \psi - \frac{z}{r \sin \chi}\right)}^{\psi} [F(z - r \sin \chi (\sin \psi - \sin t)) \sin \chi \cos t \\ + l\gamma F'(z - r \sin \chi (\sin \psi - \sin t))] e^{(r/l)t} dt \\ - \left[El \left(\sin \chi \sin \psi - \frac{z}{r} \right) - A \right] \exp \left[-\frac{r}{l} \left(\psi - \sin^{-1} \left(\sin \psi - \frac{z}{r \sin \chi} \right) \right) \right], \end{aligned}$$

when $-\frac{1}{2}\pi < \psi < \frac{3}{2}\pi$

II: $z > r \sin \chi (1 + \sin \psi)$, $d - z > r \sin \chi (1 - \sin \psi)$

$$\phi = El \sin \chi \sin \psi + \Phi_2,$$

where

$$\begin{aligned} \Phi_2 = r e^{-(r/l)\psi} \int_{-\frac{1}{2}\pi}^{\psi} [F(z - r \sin \chi (\sin \psi - \sin t)) \sin \chi \cos t + l\gamma F'(z - r \sin \chi (\sin \psi - \sin t))] e^{(r/l)t} dt \\ + \frac{r e^{-(r/l)\psi}}{e^{(r/l)2\pi} - 1} \int_{-\frac{1}{2}\pi}^{\frac{3}{2}\pi} [F(z - r \sin \chi (\sin \psi - \sin t)) \sin \chi \cos t \\ + l\gamma F'(z - r \sin \chi (\sin \psi - \sin t))] e^{(r/l)t} dt, \end{aligned}$$

when $-\frac{1}{2}\pi < \psi < \frac{3}{2}\pi$

III: $z < r \sin \chi (1 + \sin \psi)$, $d - z < r \sin \chi (1 - \sin \psi)$ (this domain exists only when $d < 2r$)

$$\phi = El \sin \chi \sin \psi + \Phi_3,$$

where $\Phi_3 = \Phi_1$, when

$$\sin^{-1} \left(\frac{z}{r \sin \chi} - 1 \right) < \psi < \pi + \sin^{-1} \left(\frac{d - z}{r \sin \chi} - 1 \right),$$

and $\Phi_3 = \Phi_4$, when

$$\pi + \sin^{-1} \left(\frac{d - z}{r \sin \chi} - 1 \right) < \psi < \frac{3\pi}{2} \quad \text{and} \quad -\frac{\pi}{2} < \psi < \sin^{-1} \left(\frac{z}{r \sin \chi} - 1 \right).$$

IV: $z > r \sin \chi (1 + \sin \psi)$, $d - z < r \sin \chi (1 - \sin \psi)$

$$\phi = El \sin \chi \sin \psi + \Phi_4,$$

where

$$\begin{aligned} \Phi_4 = r e^{-(r/l)(\psi+2\pi)} \int_{\pi - \sin^{-1}\left(\sin \psi + \frac{d-z}{r \sin \chi}\right)}^{\psi+2\pi} [F(z - r \sin \chi (\sin \psi - \sin t)) \sin \chi \cos t \\ + l\gamma F'(z - r \sin \chi (\sin \psi - \sin t))] e^{(r/l)t} dt \\ - \left[A + El \left(\sin \chi \sin \psi + \frac{d-z}{r} \right) \right] \exp \left[-\frac{r}{l} \left(\psi + \pi + \sin^{-1} \left(\sin \psi + \frac{d-z}{r \sin \chi} \right) \right) \right] \end{aligned}$$

when $-\frac{1}{2}\pi < \psi < \frac{1}{2}\pi$, and

$$\begin{aligned} \Phi_4 = r e^{-(r/l)\psi} \int_{\pi - \sin^{-1}\left(\sin \psi + \frac{d-z}{r \sin \chi}\right)}^{\psi} [F(z - r \sin \chi (\sin \psi - \sin t)) \sin \chi \cos t \\ + l\gamma F'(z - r \sin \chi (\sin \psi - \sin t))] e^{(r/l)t} dt \\ - \left[A + El \left(\sin \chi \sin \psi + \frac{d-z}{r} \right) \right] \exp \left[-\frac{r}{l} \left(\psi - \pi + \sin^{-1} \left(\sin \psi + \frac{d-z}{r \sin \chi} \right) \right) \right], \end{aligned}$$

when $\frac{1}{2}\pi < \psi < \frac{3}{2}\pi$.

function would be obtained from the analytic expression of the vanishing everywhere of the component of the current density in the direction Oz .

This, however, leads to an integral equation involving multiple integrals which examination shows to be intractable without very prolonged numerical work. In the later work we have therefore been forced to approximate by the use of the bulk Hall field throughout which leads to an over-estimation of the conductivity in general.

However, it has been possible to evaluate approximately the Hall field and its variation across the film in the case of small magnetic fields as follows.

We expand the Hall field and the distribution function in powers of $1/r$, and determine the coefficient of $1/r$ in the Hall field. We denote the complete Hall field by

$$\frac{1}{r}F_1(z) + \frac{1}{r^3}F_3(z) + \dots$$

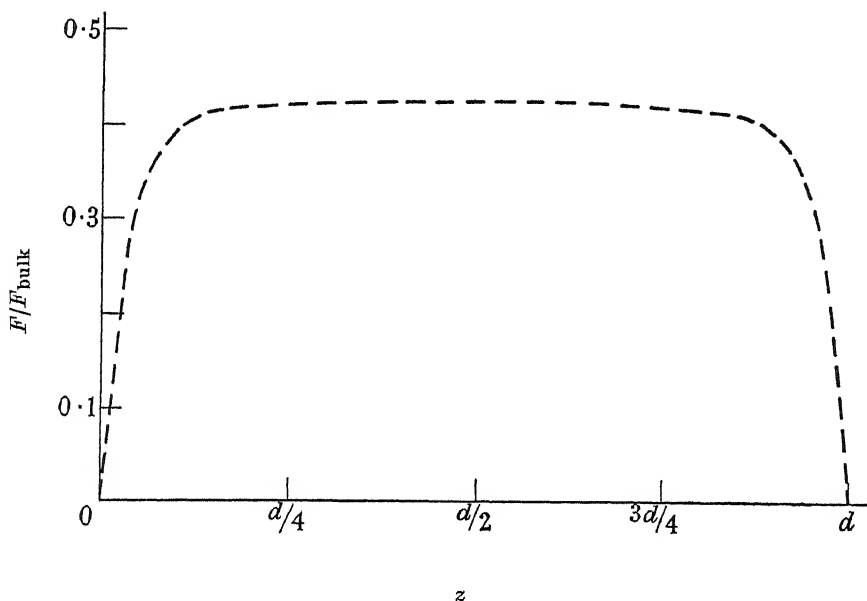


FIGURE 3. Variation of Hall field, F , across a thin film of thickness $d = 0.64 \times$ mean free path, assuming diffuse scattering at the boundaries.

(remembering that it is an odd function of H), and arrive at an implicit equation for F_1 neglecting the curvature of the orbits. In dimensionless form this reads

$$\int_0^\xi G(\eta) \alpha(\xi - \eta) d\eta + \int_\xi^{d_1} G(\eta) \alpha(\eta - \xi) d\eta + \int_0^\xi \gamma G'(\eta) \beta(\xi - \eta) d\eta - \int_\xi^{d_1} \gamma G'(\eta) \beta(\eta - \xi) d\eta + (\lambda + \frac{1}{6}\xi) \alpha(\xi) + (\lambda + \frac{1}{6}(d_1 - \xi)) \alpha(d_1 - \xi) - \frac{1}{3}(2 - e^{-\xi} - e^{-(d_1 - \xi)}) = 0,$$

where $l\xi = z$, $ElG(\eta) = F_1(l\eta)$, $ld_1 = d$, $\frac{\lambda El^2}{r} = A$, $\alpha(x) = \int_1^\infty e^{-xt} \frac{dt}{t^3}$, $\beta(x) = \int_1^\infty e^{-xt} \frac{dt}{t^2}$.

The equation was solved by numerical methods. Since the Hall field assumes, at most, its bulk value El/r , $\gamma G'(\eta)$ will be negligible except possibly for *very* small values of η . We have therefore neglected the third and fourth integrals in the calculation which we have carried out for $d_1 = 0.64$ (corresponding to $l \approx 45\mu$, $d \approx 29\mu$). The calculated value for λ is 0.1, and the variation of the Hall field is shown in figure 3.

An estimate made for the case also of $d_1 = 2(d = 90\mu)$ showed that the Hall field then attains a value ~ 0.7 of the bulk value.

The rapid rise of the Hall field as one leaves an edge appears somewhat remarkable to us, and at present we cannot suggest any obvious physical reason for this behaviour, which one might expect could be expressed in terms of some 'characteristic depth'.

3.23. Determination of the electrical conductivity

In order to evaluate the electric current, and hence the conductivity, we require

$$J_x = \frac{e}{d} \int_0^d dz \iiint N v_x dv_x dv_y dv_z, \quad (14)$$

where the expressions for the distribution function appropriate to the various regions of velocity space are set out in table 2.

In view of the rapid establishment of the Hall field at an edge we have made the simplifying approximation of a constant Hall field across the film. In our main calculation we have in fact assumed that this has the bulk value El/r throughout, but we add some discussion later about the error involved. Concomitantly we neglect the terms involving γ in table 2; these terms (which are small anyway) are appropriate to the rise of the Hall field at the edge, and such error as may arise will be in the opposite sense to that from the assumption of a uniform Hall field. We also omit the constant A appearing in table 2; for small fields (large r) the contribution from the appropriate term is negligible, while for large fields ($r \rightarrow 0$) the regions of the film where the term contributes are very small.

One would naturally wish to avoid all approximations but these appear to be the minimum necessary to permit reasonable computation.

In *extenso* equation (14) leads, then, to

$$\begin{aligned} \frac{J_x}{J_{\text{bulk}}} = 1 + \frac{3}{El d \pi} & \left\{ \int_{\chi=0}^{\alpha} \int_{\psi=-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \left(\int_{z=0}^{r \sin \chi (1 + \sin \psi)} \Phi_1 dz \right. \right. \\ & + \frac{1}{2} \int_{z=r \sin \chi (1 + \sin \psi)}^{2r \sin \alpha - r \sin \chi (1 - \sin \psi)} \Phi_2 dz \Big) \sin^2 \chi \sin \psi d\chi d\psi \\ & + \int_{\chi=\alpha}^{\frac{1}{2}\pi} \left(\int_{\psi=-\frac{1}{2}\pi}^{\sin^{-1} \left(\frac{2 \sin \alpha}{\sin \chi} - 1 \right)} \int_{z=0}^{r \sin \chi (1 + \sin \psi)} \Phi_1 dz \sin \psi d\psi \right. \\ & + \int_{\psi=\sin^{-1} \left(\frac{2 \sin \alpha}{\sin \chi} - 1 \right)}^{\frac{1}{2}\pi} \int_{z=0}^{2r \sin \alpha} \Phi_1 dz \sin \psi d\psi \\ & \left. \left. + \int_{\psi=\frac{1}{2}\pi}^{\pi + \sin^{-1} \left(\frac{2 \sin \alpha}{\sin \chi} - 1 \right)} \int_{z=0}^{2r \sin \alpha - r \sin \chi (1 - \sin \psi)} \Phi_1 dz \sin \psi d\psi \right) \sin^2 \chi d\chi \right\} \quad (15) \end{aligned}$$

for the case of a film such that $d < 2r$, and we write $d = 2r \sin \alpha$, for convenience.

Also

$$\begin{aligned} \frac{J_x}{J_{\text{bulk}}} = 1 + \frac{3}{El d \pi} & \left\{ \int_{\chi=0}^{\frac{1}{2}\pi} \int_{\psi=-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \int_{z=0}^{r \sin \chi (1 + \sin \psi)} \Phi_1 dz \sin \psi \sin^2 \chi d\psi d\chi \right. \\ & \left. + \frac{1}{2} \int_{\chi=0}^{\frac{1}{2}\pi} \int_{\psi=-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \int_{z=r \sin \chi (1 + \sin \psi)}^{d - r \sin \chi (1 - \sin \psi)} \Phi_2 dz \sin \psi \sin^2 \chi d\psi d\chi \right\} \quad (16) \end{aligned}$$

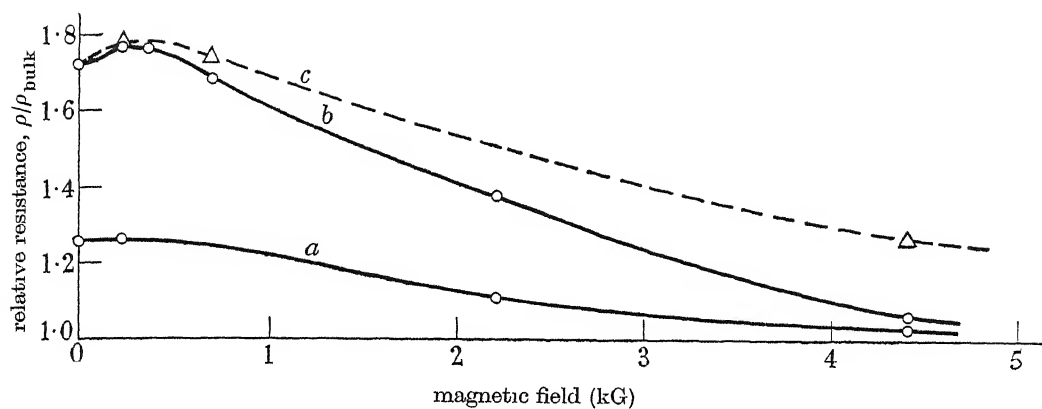


FIGURE 4. Theoretical curves for resistance of thin sodium films *a*, 58 μ thickness; *b*, 29 μ thickness (Hall field assumed bulk value throughout); *c*, 29 μ thickness (estimated correct curve—see text).

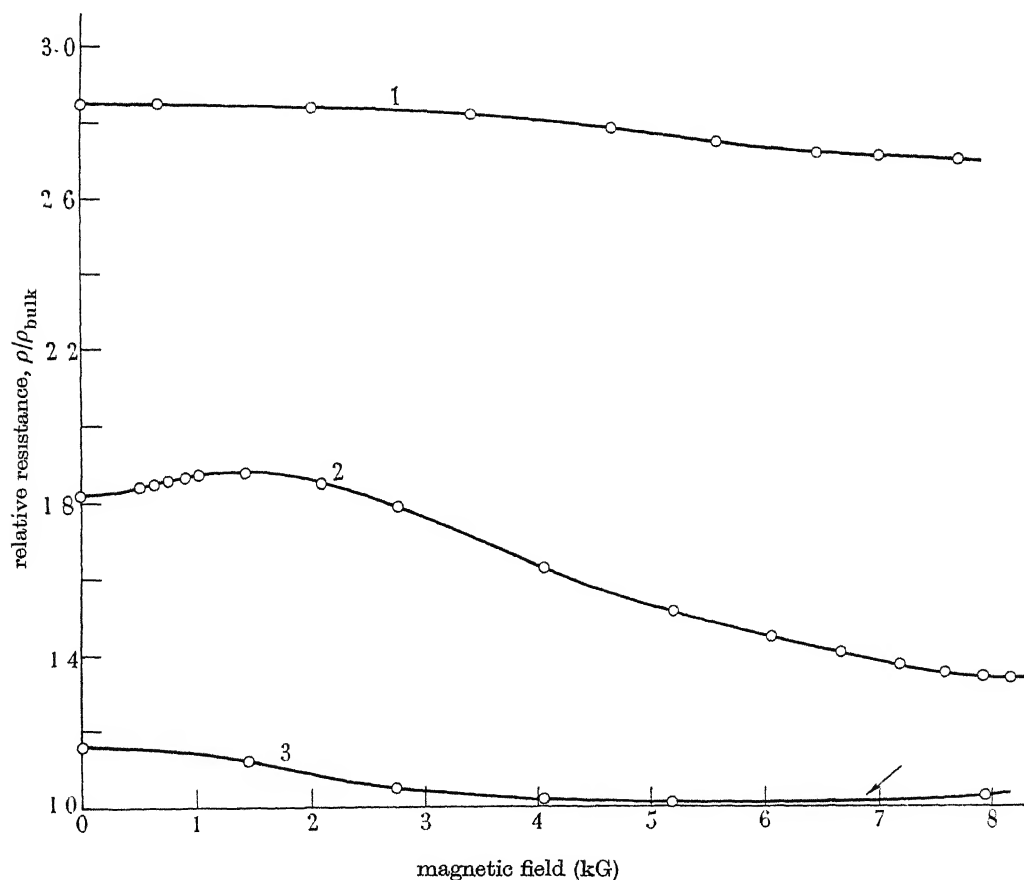


FIGURE 5. Experimental results on sodium cylindrical specimens under a transverse magnetic field. Diameters. 1, 20 μ ; 2, 30 μ ; 3, 66 μ . The arrow indicates the appearance of inherent magneto-resistance. All results at $T = 4.2^\circ \text{K}$ ($l \approx 44 \mu$).

for the case of a film such that $d > 2r$. These expressions do not appear to involve Φ_4 , but in writing them down use is made of the symmetry between Φ_1 and Φ_4 , and, in fact, the correct result for the conductivity is obtained by doubling the contribution from Φ_1 . Equation (15) can ultimately be reduced with the approximations mentioned above to a sum of irreducible double and single integrals. These we have evaluated numerically. Equation (16) allows direct integration to give

$$\frac{J_x}{J_{\text{bulk}}} = 1 - \frac{9lr^2}{32d(r^2 + l^2)} \left\{ \frac{4}{3} + \frac{2l^2}{l^2 + r^2} (1 - e^{-(2r/l)\pi}) + \frac{(26l^2 + 24r^2)l^2}{(l^2 + 4r^2)(9l^2 + 4r^2)} (1 + e^{-(2r/l)\pi}) \right\}. \quad (17)$$

These results are plotted in figure 4, for films with $d \approx 29 \mu$, 58μ , assuming $l = 45 \mu$ and $v \approx 1.1 \times 10^8$ cm./sec. (cf. Mott & Jones 1936, p. 268) These values were chosen as being close to particular cases that had been investigated experimentally, and the corresponding experimental curves appear on figure 5, together with other data on figure 6.

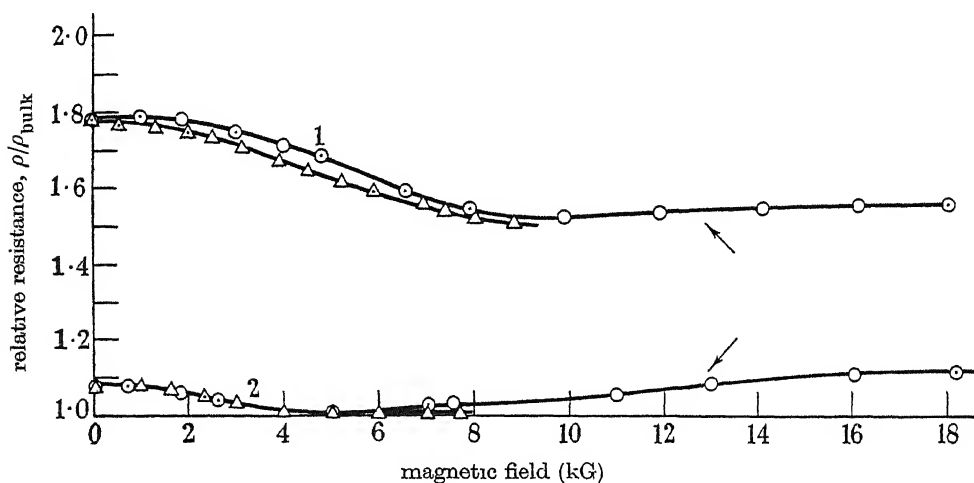


FIGURE 6. Experimental results on sodium cylindrical specimens under transverse and longitudinal magnetic fields. Diameter of specimens: 1, 30μ ; 2, 66μ . \circ , transverse field; \triangle , longitudinal field. All results at $\sim 11^\circ$ K ($l \approx 37 \mu$). The arrows indicate the appearance of inherent magneto-resistance.

4. DISCUSSION AND CONCLUSION

Direct quantitative comparison of theory and experiment is not possible for two primary reasons. First, the experimental data were obtained on cylindrical specimens and, while the theory may be extended formally on the same lines to a square wire, we have not attempted to evaluate the conductivity numerically in that case on account of the complexity of the integrals involved. Secondly, it is clear from §2 that a degree of elastic scattering is present in all specimens while diffuse scattering is assumed in the theory. Consequently no attempt has been made to plot theory and experiment on the same graph. It is clear that the qualitative features of the experiments are well predicted by the theory, in particular, the rapid reduction of the initial 'hump' as the specimen size is increased. The very much quicker descent

of the resistance with magnetic field given by the theory in the case of the thinner specimen is due considerably to the geometry, since the constricting effect of the cylindrical structure will obviously become stronger as the size is reduced; this is clearly evidenced by the experimental data on the 20μ cylinder in figure 5. The fractional rate of decrease in the relatively large 66μ specimen shows in fact much better quantitative agreement between theory and experiment.

A further problem is the value of the Hall field, as mentioned earlier. For very small fields the value chosen has little effect, as may be seen from figure 5, where we have also evaluated the initial variation of resistance with a Hall field = 0.4 times the bulk value as determined in §3.22. As the field increases the deviation from the bulk value causes the resistance to decrease more slowly, we have roughly estimated that at 4000 G the Hall field should have attained ~ 0.8 time the bulk value and on this basis computed a value for the resistance as shown. This value, coupled with the 'corrected' part of the curve for small fields, gives one a fair idea of the overall correct curve for a film specimen.

It should be observed that the Hall field and its variation will in fact alter with the law of scattering. In particular it is interesting to notice that even with perfectly elastic scattering the conductivity in a thin film will show a dependence on a transverse magnetic field applied in the plane of the film because of the non-uniformity of the Hall field.

It appears desirable to make experiments in the future on films of sodium in order to check the theory more closely. At the same time the effect of a rotation of the magnetic field through 90° perpendicular to the electric field could be checked against Sondheimer's theory.

Sodium (and possibly pure lithium and potassium) appears to be the only metal suitable for these experiments in view of the higher inherent bulk magneto-resistance of other metals, even the heavier alkali metals (cf. MacDonald 1950). Experiments were in fact carried out on thin foils of tin, silver and gold. A small difference in the correct sense between the two orientations of the magnetic field was observed with tin, but the inherent magneto-resistance was the dominating factor in these metals, and also with silver and gold the residual resistance was rather high resulting in a short mean free path. Thus in the case of silver the mean free path at $\sim 4^\circ$ K was $\sim 7\mu$, and the magneto-resistance increase $\sim 13\%$ for about 8 kG.

A simple approximate mean free-path theory was developed for the change of resistance in a cylindrical wire under a longitudinal magnetic field which showed reasonable agreement with experiments. While this paper was being prepared, however, we heard of Mr Chambers's detailed work on this problem, and we have therefore not considered it worth while to include the analysis here, particularly as Mr Chambers has also carried out a collateral set of experiments in a longitudinal magnetic field on our 30μ sodium specimen.

We should like to thank Professor M. H. L. Pryce and Mr J. L. Olsen for helpful discussions in the early stages of the investigation. This work was carried out during the tenure of an I.C.I. Research Fellowship by one of us (D.K.C. MacD.).

REFERENCES

- Andrew, E. R. 1949 *Proc. Phys. Soc.* **62**, 77.
 Chambers, R. G. 1950 *Proc. Roy. Soc. A*, **202**, 378.
 Dingle, R. B. 1950 *Proc. Roy. Soc. A*, **201**, 545.
 Fuchs, K. 1938 *Proc. Camb. Phil. Soc.* **34**, 100.
 Lovell, A. C. B. 1936 *Proc. Roy. Soc. A*, **157**, 311.
 MacDonald, D. K. C. 1947 *J. Sci. Instrum.* **24**, 232.
 MacDonald, D. K. C. 1950 *Proc. Phys. Soc. A*, **63**, 290.
 MacDonald, D. K. C. & Mendelssohn, K. 1948 *Nature*, **161**, 972.
 MacDonald, D. K. C. & Mendelssohn, K. 1950 *Proc. Roy. Soc. A*, **202**, 103.
 Mott, N. F. & Jones, H. 1936 *Theory of properties of metals and alloys*.
 Oxford University Press.
 Nordheim, L. 1931 *Ann. Phys., Lpz.*, **9**, 607.
 Nordheim, L. 1934 *Act. Sci. et Ind.* no. 131. Paris: Hermann.
 Sarginson, K. & MacDonald, D. K. C. 1949 *Nature*, **164**, 921.
 Simon, F. E. 1936 *Act. 7th Int. Congr. Refrig.* **1**, 367.
 Sondheimer, E. H. 1949 *Nature*, **164**, 920.
 Stone, I. 1898 *Phys. Rev.* **6**, 1.
 Thomson, J. J. 1901 *Proc. Camb. Phil. Soc.* **11**, 120.
 Tj̄peica, S. 1935 *Ann. Phys., Lpz.*, **22**, 129.
 Wilson, A. H. 1936 *The theory of metals*, p. 161. Cambridge University Press.

Liquid helium films

II. The flow of the film

BY K. R. ATKINS

The Royal Society Mond Laboratory, University of Cambridge

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 29 March 1950)

The rate of flow of the helium II film has been found to vary slightly with the height of the film and the pressure head. Most of the variation can be ascribed to changes in the thickness of the film, but a small residual variation at small pressure heads suggests that at least part of the flow is subject to frictional retardation. Measurements of the damping of the film oscillations enable a small upper limit to be put on the frictional retardation of the remaining part, but a similar method applied to flow through wide capillaries indicates that in this case most of the flow is opposed by frictional forces. The results are discussed in conjunction with the measurements of Allen & Misener on flow through capillaries and an attempt made to formulate some of the general features of the flow of the superfluid component.

1. INTRODUCTION

A very satisfactory qualitative explanation of the anomalous transport properties of liquid helium II is provided by the two-fluid theories of Tisza (1938, 1947) and Landau (1941), who assume that the liquid is a mixture of a normal component and a superfluid component. The normal component possesses a normal viscosity and carries most or all of the heat content of the liquid, whereas the superfluid component flows with appreciable velocities through the very narrowest channels

but carries very little heat with it. Further development of the theory requires a more precise formulation of the hydrodynamical equations applicable to the flow of the two components, and the purpose of this paper is to describe some experiments on the flow of the liquid helium film and to discuss the information they provide about the hydrodynamics of the superfluid component. The particular virtue of the film for this purpose is that its thickness is only about 2×10^{-6} cm., which is considerably smaller than any channel which can conveniently be made artificially and is certainly small enough to inhibit the flow of the normal component completely.

The flow of the film was first studied by Daunt & Mendelssohn (1939), who found that the rate of flow was practically independent of the pressure head and therefore postulated the existence of a critical rate of flow attained by the film under all circumstances and varying only with the temperature. In a later experiment (1946) they demonstrated that a velocity almost as great as the critical velocity could be realized under a pressure head not greater than 0.1 mm. Bijl, de Boer & Michels (1941) consider that this critical velocity v_c is determined by a quantum relationship of the type

$$mv_c d = \alpha h, \quad (1)$$

in which m is the mass of the helium atom, d is the thickness of the film, h is Planck's constant and α is a constant not very different from unity. This predicts that $v_c d$, which is the rate of flow of the film, should be independent of d . Velocities considerably greater than v_c were found by the present author (Atkins 1948) and by van den Berg & de Haas (1949), but have been shown by Bowers & Mendelssohn (1949) to be due to a granular deposit of some solidified gas on the surface holding the film and are therefore not fundamental. (In the experiments to be described in this paper great care was taken to avoid these contamination effects. The details have been discussed in part I (Atkins 1950). Gorter & Mellink (1949) have made an analysis of the thermal conductivity measurements in narrow channels and have challenged the concept of critical velocity, substituting for it a mutual frictional force between the normal and superfluid components. The three outstanding problems which will be borne in mind throughout this paper are therefore: (1) Is there anything which can reasonably be described as a characteristic velocity? (2) How does the rate of flow vary with the thickness of the film? (3) Is the flow of the superfluid component truly frictionless or are there small dissipative forces such as the mutual friction term in the theory of Gorter & Mellink?

2. VARIATION WITH THE FILM THICKNESS

(a) *Influence of the height of the film*

During the oscillation experiments described in part I it was a simple matter to study the motion of the inner meniscus before the start of the oscillations, and thus to make simultaneous measurements of the thickness of the film and its rate of flow. The rate of flow was found to vary slightly with both the pressure head and the height of the film, but in a complicated manner. The pressure-head effects will be discussed in more detail in a later section, and at present we shall confine ourselves to the statement that, when the vessel was filling and the pressure head was greater than

a few millimetres, the rate of filling was completely independent of the pressure head (figure 4). Under these circumstances, therefore, it was permissible to talk of a characteristic rate of flow σ_c and a characteristic velocity $v_c = \sigma_c/d$. When adequate precautions were taken to prevent the formation of solid deposits on the surface holding the film, this characteristic rate of flow had the same order of magnitude as in Daunt & Mendelssohn's (1939) experiments, but varied with the height of the

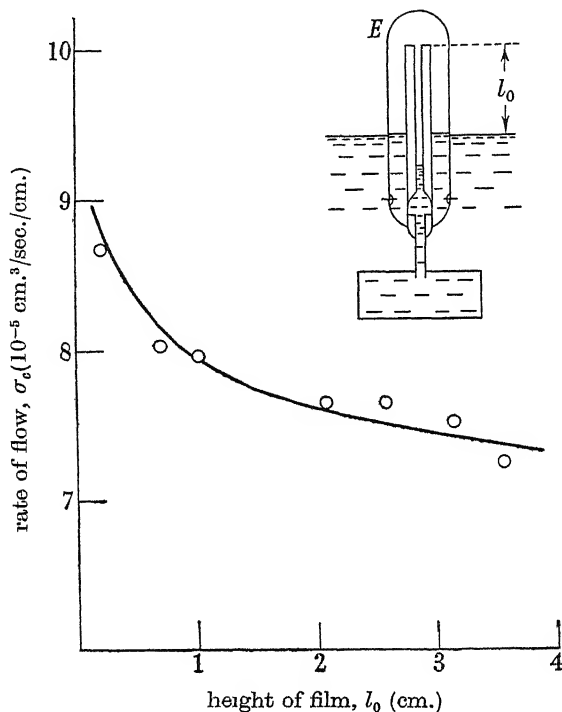


FIGURE 1. Variation of the rate of filling with the height of the film.
Temperature = 1.47° K.

film l_0 in the manner shown in figure 1. This type of behaviour was observed in all the experiments, but the results of figure 1 were obtained with the apparatus which was referred to as vessel no. 2 in part I, and the exact details of its construction are shown in the inset to figure 1. It was similar in design to vessel no. 1 (see figure 3 of part I) but the down-going wide tube X had been replaced by the envelope E , and as a result the film height l_0 was a more definite and more accurately measurable length.

In their original experiments Daunt & Mendelssohn (1939) observed an increase in the rate of flow when one level was near the rim of their beaker, but they suggested that it was due to surface-tension effects. In actual fact the surface-tension rise of a liquid helium meniscus near a wall is about 0.7 mm., whereas in figure 1 there is a significant variation of the rate of flow at film heights of several centimetres, and this must be regarded as a genuine film effect. It is probably a consequence of the variation of film thickness with height which was established in part I, and, if this is so, provides information on the variation of the rate of flow with the channel width.

The oscillation results suggest that, over the range of heights from 0.5 to 5 cm., the film thickness can be represented fairly well by the equation

$$d = 2.2 \times 10^{-6} / H^{0.14},$$

in which the value of k is that appropriate to vessel no. 2, and the exponent n is a mean for this range of heights and is the same for the two vessels within the experimental error. This formula, combined with the assumption that σ_c is determined by the value of d at the rim of the vessel, leads to the curve of figure 2 for the variation of σ_c with d . The fact that σ_c is not independent of d means that the equation

$$m\sigma_c = mv_c d = \alpha h$$

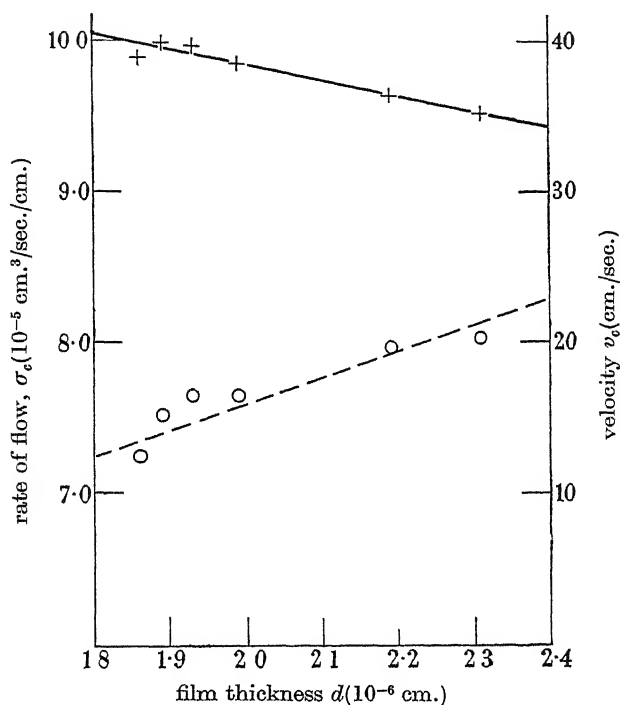


FIGURE 2 Variation of σ_c and v_c with d for the film. O, σ_c ; +, v_c .

cannot be exactly true, although, of course, the fundamental principle underlying it may still be valid. In figure 2 the characteristic velocity $v_c = \sigma_c/d$ is also shown as a function of d , and it will be seen that it is not constant either, but decreases as d increases. The values of v_c given here are considerably higher than the value of about 20 cm.sec.⁻¹ which has previously been discussed in the literature and which was based on a film thickness of 4×10^{-6} cm.

(b) Comparison with the subsurface measurements

Allen & Misener (1939) have studied the dependence of the rate of flow on the channel width for flow through glass capillaries and wire-filled tubes. Their results are in qualitative agreement with the results of the previous section, for they also found that as the channel width was increased the rate of flow increased but the mean

velocity decreased. Figure 3 includes for comparison with the present results the measurements they made at their lowest temperature (1.20°K), and at their lowest pressure gradient ($0.75\text{ dyne cm.}^{-3}$). Although the two sets of measurements cover different ranges of d , their general trend is similar. The curve of figure 3 can be divided into two parts, depending on whether d is greater than or less than about 10^{-4} cm . For large values of d , the velocity is approximately independent of d , but

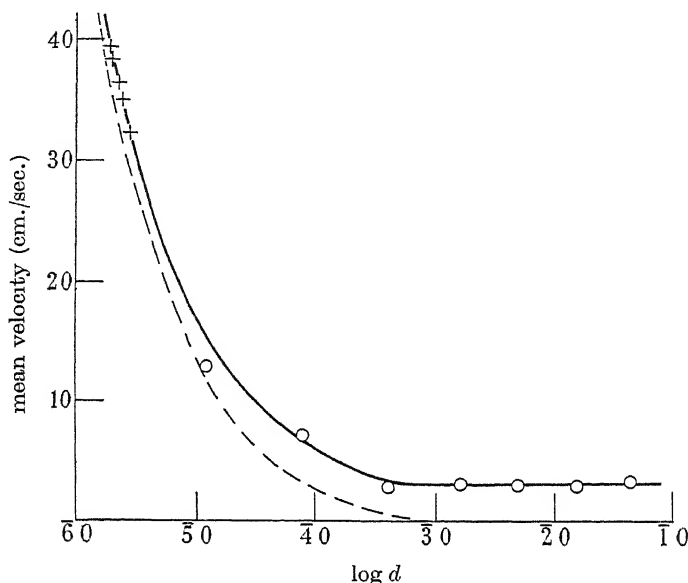


FIGURE 3. Comparison of the film results with the measurements of Allen & Misener. +, characteristic velocity v_c in the film at 1.47°K , O, mean velocity in Allen & Misener's experiment at 1.2°K under a pressure gradient of $0.75\text{ dyne cm.}^{-3}$; ---, v_1 (see § 5).

in this region there is a pronounced non-linear dependence on the pressure head. When d is less than 10^{-4} cm , the velocity increases rapidly as d decreases, but pressure-head effects become increasingly less important until in the case of the film they are almost negligible. To explain this it has frequently been suggested that the flow is the sum of a frictionless part σ_1 , and a pressure-dependent part σ_2 subject to frictional retardation. On this assumption, σ_1 predominates in very narrow channels, but as the channel width increases σ_2 becomes increasingly more important. To test this theory we shall now examine the evidence which points towards the flow of the film being truly frictionless, and in the course of the discussion some evidence will be presented which suggests that most of the flow through wider channels is not frictionless.

3. PRESSURE-HEAD EFFECTS

(a) Large pressure heads

The features now to be discussed were unmistakably present in the capillary experiments described above, but it was found more convenient to examine them in detail using a glass beaker of larger internal diameter (0.347 cm), because then the

inner meniscus moved much more slowly and its velocity could be measured more accurately. Figure 4 shows how the rate of transfer varied with the difference in height of the two liquid levels for this vessel. Confining attention for the present to level differences greater than about 0.5 cm., the most striking feature is that the rate of filling was independent of the level difference to a high degree of accuracy, and this is the justification for the use in § 2 of a critical rate of flow σ_c in accordance with the ideas of Daunt & Mendelssohn (1939). The emptying curve of figure 4 shows

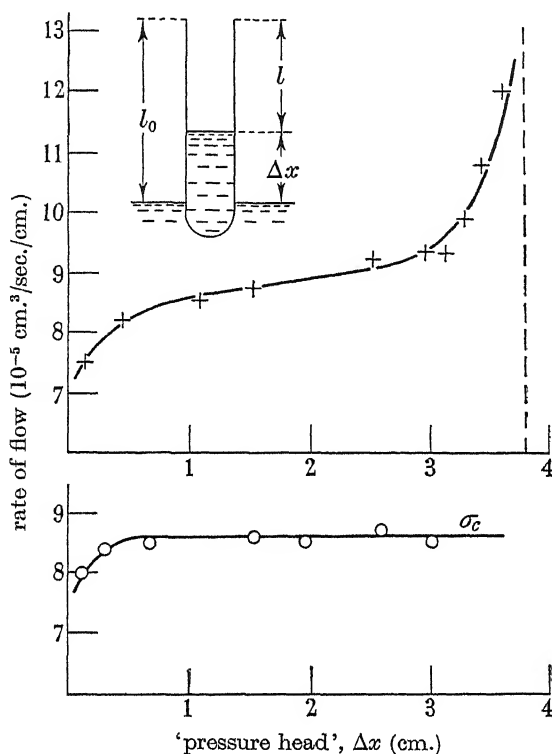


FIGURE 4 Dependence of the rate of flow of the film on the level difference.
+, emptying, $l_0 = 3.78$ cm.; O, filling, $l_0 = 0.82$ cm.

a marked dependence on the level difference, but this is probably not a true pressure-head effect but a consequence of the variation of film thickness with height. If we make the reasonable assumption that the thickness of the film at the rim of the vessel is determined by the height l of the rim above the higher of the two liquid levels, then during a filling the higher level is the outer level, which remains stationary, and so the thickness of the film remains constant. During an emptying, however, the higher level is the inner level, which is steadily falling, and so the film thickness at the rim gradually decreases as the motion progresses, thus producing an apparent dependence on the pressure head. The emptying curve of figure 4 is therefore similar in character to the curve of figure 1, but it shows much more clearly how the rate of transfer increases rapidly at small film heights where δ is increasing rapidly.

(b) Small pressure heads

Both the curves of figure 4 dip down sharply when the pressure head falls below 0.5 cm. Since this effect is small and occurs only over the last few millimetres of the motion, it can be detected only in the most careful measurements, but it was found to occur in several of the more favourable experiments using various vessels, and it is considered to be quite genuine. Moreover, there is additional evidence in its favour derived from the amplitudes of the oscillations. For a sine-wave oscillation the rate of transfer $(dx/dt)_0$ at the start of the oscillation is related to the amplitude x_0 by the equation

$$\gamma = \frac{2\pi x_0}{\tau(dx/dt)_0} = 1. \quad (2)$$

To make use of this expression a small damping correction was applied to x_0 , and the mean of a large number of observations was taken to reduce the rather large error, $\sim 10\%$, inherent in the measurement of the small quantity x_0 . Ignoring the possibility of a dip at the lower end of the curve, $(dx/dt)_0$ was obtained by extrapolating to zero pressure head the rates of transfer at large pressure heads. The values thus obtained for γ in the experiment using vessel no. 2 are shown in table 1. The mean value of γ is 0.93 ± 0.02 , which is significantly less than unity. This can be explained if the value of (dx/dt) just before the start of the oscillations is slightly less than the value obtained by extrapolation from large pressure heads, implying the existence of the dip.

TABLE 1

film height l_0 (cm.)	x_0 (divisions)	γ
3.25	3.12	0.91
3.57	3.25	0.85
3.16	3.25	0.87
2.60	3.12	0.92
2.10	2.62	0.89
1.03	1.75	1.01
0.61	1.00	0.90
0.99	2.00	1.10
1.95	2.25	0.93
2.60	2.62	0.91

The behaviour of the flow at very small pressure heads is of fundamental importance, for if part of the flow is frictionless then this part should be able to exist under zero pressure head. In fact, in a steady state, the pressure head is a direct measure of the frictional forces opposing the flow, and a graph such as figure 4 showing how the pressure head varies with the rate of flow is exactly equivalent to a graph showing how the frictional force varies with the velocity. The important possibilities for the nature of the frictional force are illustrated in figure 5. Curve 1 refers to ordinary viscous forces which give a linear variation of flow with pressure head. The theory of Gorter & Mellink (1949) postulates a mutual force between ρ_s and ρ_n approximately proportional to the cube of the relative velocity as in curve 2. Curve 3 corresponds to the simplest form of the critical velocity hypothesis, which assumes that the friction is zero below v_c , but becomes infinite at v_c , so that v_c can

never be exceeded. Curve 4 represents a mixed flow composed of a frictionless part which can have velocities up to v_c and a pressure-dependent part which can raise the total velocity above v_c but is subject to frictional retardation. Curve 4 seems to provide the best explanation of the film results, but we must not exclude the possibility, represented by curve 5, that there is a small amount of friction at all velocities. Daunt & Mendelssohn (1946) have shown that the film can have a velocity of $0.9v_c$ under a pressure head not greater than 10^{-2} cm., but this result can be fitted on to the curves of figure 4 and still allow the possibility that the velocity falls to zero in the range of pressure heads below this.

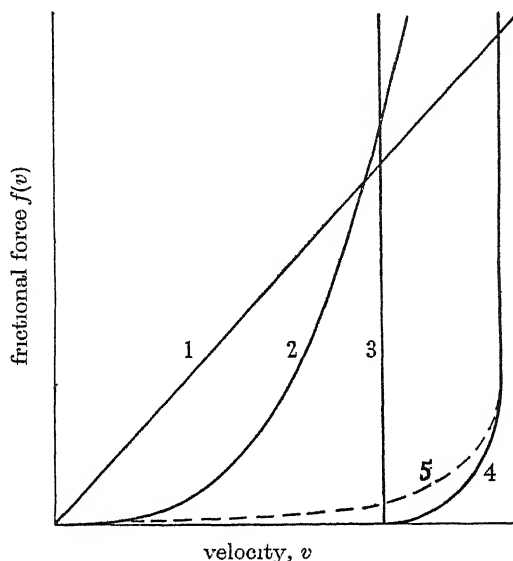


FIGURE 5. Dependence of the frictional forces on the velocity. 1, ordinary viscous forces; 2, Gorter-Mellink theory; 3, simple critical velocity; 4, mixed flow; 5, flow which is never completely frictionless.

(c) The oscillations

The equation of motion of the film was derived in §2(b) of part I. Introducing an extra frictional term f varying in an unspecified way with the velocity, the equation for the height x of the inner level above the outer level becomes

$$\frac{d^2x}{dt^2} + f\left(\frac{dx}{dt}\right) + w^2x = 0. \quad (3)$$

In a steady state, when d^2x/dt^2 is zero or negligible, this reduces to

$$x = -\frac{1}{w^2}f\left(\frac{dx}{dt}\right), \quad (4)$$

which determines the way in which the pressure head x varies with the rate of flow dx/dt . f is therefore the function which has been discussed in the preceding section in connexion with figure 5. As x approaches zero the acceleration d^2x/dt^2 becomes important and the oscillations commence. The term f then produces a damping of the oscillations, and information about f at small values of v can be obtained from

a study of this damping. The point which immediately emerges from such considerations is that the oscillations could not exist at all unless the frictional forces were extremely small. Assuming, for simplicity, to obtain an order of magnitude result, that $f(dx/dt)$ has the form $2k dx/dt$, then the solution of equation (3) is

$$x = x_0 e^{-kt} \sin \sqrt{(w^2 - k^2)} t.$$

The value of k which would damp down the oscillations in one period would be so small that the characteristic velocity v_c would be maintained by a pressure head of only 10^{-2} cm. Actually, it is clear from figure 4 of part I that the damping is much

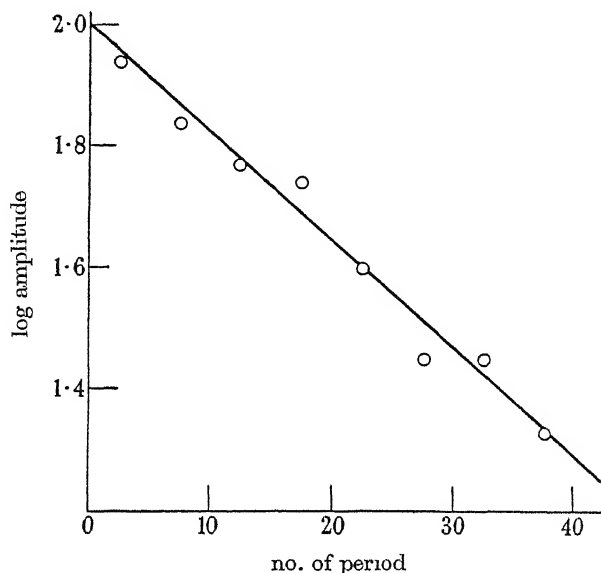


FIGURE 6. Damping of the oscillations. Temperature = 1.10° K. Period = 27.8 sec. Film height = 4.2 cm. Logarithmic decrement = 0.019.

less than this. Figure 6 shows the results of some measurements which were made with vessel no. 2 at the lowest temperature attainable. From the observed logarithmic decrement in this case the treatment given above enables us to deduce that a rate of flow σ_c would be produced by a pressure head of only 2×10^{-4} cm. The linearity of the curve in figure 6 suggests that the frictional forces are proportional to the velocity, but it should be remembered that the damping of the oscillations may be due to some secondary factor rather than to fundamental frictional forces opposing the flow, and so the argument merely puts an upper limit to such forces. We may say, therefore, that the damping of the oscillations proves that there is very little friction associated with the flow of the film, and leaves open the possibility that the flow is completely frictionless.

(d) *Wide capillaries*

If there is a critical velocity in wide capillaries it is clear from figure 3 that it cannot be greater than 3 cm. sec.^{-1} , but, as the flow is so pressure-dependent in these capillaries, the true critical velocity may be much less than this. In circumstances like this the oscillations provide a very good method of distinguishing between the

frictionless and non-frictionless parts of the flow, for, as was shown in the last section, oscillations can only result from that part of the flow which is frictionless or, at the most, subject to extremely small frictional forces. If the frictionless part of the flow has a velocity v_1 , then the amplitude of the oscillations is

$$x_0 = v_1 \sqrt{\left(\frac{\rho_s}{\rho} \frac{Sl}{\pi R^2 g} \right)}, \quad (5)$$

in which S is the cross-sectional area available for flow, l is the length of the capillary and R is the radius of the reservoir, and so v_1 can be estimated from the observed amplitude. This method was used to investigate the flow through a capillary 8 cm. long with an internal diameter of 0.6 mm., using a reservoir of radius 3 mm. With this arrangement the period of the oscillations would have been 6 sec. Actually, at 1.23° K, no oscillations with an amplitude greater than 2×10^{-3} cm. were observed. This implies that v_1 in this capillary was less than 0.2 cm. sec.⁻¹, and therefore most of the flow in capillaries of this width must be pressure-dependent. Experiments are in hand to apply this method to channel widths in the whole range from the film width upwards and thus to determine the relationship between v_1 and d .

4. THERMAL EFFECTS

Liquid helium flow experiments are subject to two types of thermal disturbance, arising, first, from the heat defect of the liquid which flows and, secondly, from the radiant heat incident upon the apparatus. We shall now discuss the relevance of these factors to the present experiments. The flow of film into the vessel results in a supply of cold to the inside which can be calculated from the observed rate of flow and the heat defect of the moving part of the film as measured by Kapitza (1942). A steady state is soon reached in which the temperature inside is slightly lower than outside and, in the type of apparatus shown in figure 3 of part I, the cold brought in by the film is compensated by heat conducted through the copper foil provided for that purpose. The main thermal resistance is that of the copper-liquid boundary and has been measured by Kapitza (1941), so that the temperature difference is easily calculable and turns out to be of the order of magnitude of 10^{-6} ° K at 1.47° K. This temperature difference would produce a thermomechanical pressure head of only 2×10^{-3} cm. of helium, which is so small compared with the hydrostatic pressure heads used that it can be entirely neglected. In an apparatus without a copper foil, such as that shown in figure 4, the cold brought over by the film is dissipated by a distillation process in which liquid evaporates from the bath and condenses inside the vessel, giving up its latent heat as it does so. Since the latent heat is about 80 times as large as the heat defect of the superfluid component at 1.47° K, the transfer by distillation need be only 1.2 % of the surface transfer. Assuming the vaporization coefficient to be unity, the temperature difference necessary to produce this distillation may be shown to be of the order of 10^{-6} ° K for the vessel of figure 4 and is again negligible (for the theory of evaporation rates, see, for example, Knudsen 1934).

Radiant heat incident upon the apparatus raises the inside temperature above that of the outside, producing a thermomechanical effect and also a steady evaporation

of liquid out of the vessel. Since the thermomechanical effect was not observable, the steady temperature difference could not have been greater than 10^{-5}°K , but there was probably an appreciable rate of evaporation which increased the measured rate of emptying and decreased the measured rate of filling. Because of the complications apparent from figure 4, it was not easy to make an exact comparison between the rates of emptying and filling at a constant film height, but it appeared that the rate of emptying was consistently greater than the rate of filling by about 10 %. The rates of flow and velocities given in figures 1, 2 and 3 are therefore probably too small by about 5 %. This does not affect any of the conclusions drawn from these results.

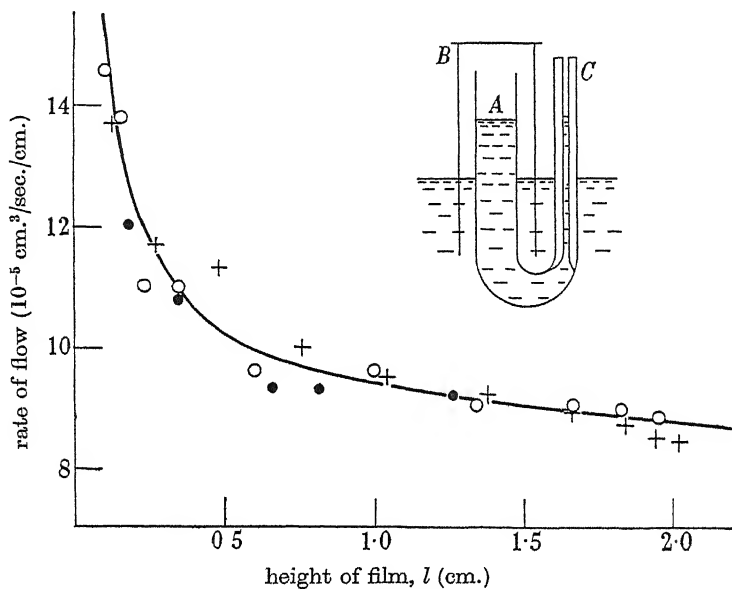


FIGURE 7. Transfer rates of a film protected from radiation.
+, protected film; O, ●, unprotected film.

Another possibility is that the radiation directly incident upon the film influences the situation, either by changing the film thickness or by producing an evaporation from the surface of the film. To test this point, an experiment was performed with a film protected from radiation in the manner shown in the inset to figure 7. The U-shaped vessel emptied through the film formed on the walls of the wider limb A, which was completely surrounded by the metal radiation shield B. The other limb C was a monitoring capillary which enabled the level of the liquid in the U-tube to be observed from outside. As figure 7 shows, the order of magnitude of the transfer rate and its variation with film height were the same as in two similar experiments in which the film was not protected from radiation. (This figure also gives some idea of the reproducibility of the transfer rates.) It follows that the high rates of transfer observed by van den Berg & de Haas (1949) must have been owing to solid deposits and not to the fact that their film was protected from radiation.

5. DISCUSSION

The results of Allen & Misener (1939) at 1.2° K refer to the flow of practically pure superfluid helium, since the viscous flow of the small amount of normal component present at this temperature can be shown to be negligible. Moreover, the thermal effects which influenced their measurements at higher temperatures were not important at this low temperature. Their results at this temperature will therefore be combined with the results described above in an attempt to formulate some of the outstanding features of the flow of the superfluid component. To simplify the discussion, the mean velocity of flow in Allen & Misener's experiments will be expressed in the form

$$\begin{aligned} v &= v_1 + v_2 + v_3 \\ &= \sigma_1/d + v_2 (\text{grad } p) + v_3 (\text{grad } p, d), \end{aligned}$$

but it must be borne in mind that this formula may be purely empirical and have no theoretical foundation. v_3 is important in the widest capillaries, v_2 in capillaries of intermediate width and v_1 in the narrowest capillaries and in the film.

v_3 varies with both the channel width d and the pressure gradient. Since the flow is approximately proportional to the square root of the pressure gradient in wide capillaries, v_3 may be due to turbulent flow of the superfluid component. Another possible mechanism arises from the fact that the velocity with which the normal component flows through the tube is quite appreciable, even though its rate of flow in g./sec. is small because of its small density. Therefore, if there is a mutual force between the two components, the normal component will tend to drag the superfluid component along with it. However, whatever the exact nature of the mechanism, it may also be important in viscosity experiments involving the damping of the oscillations of a disk, and the apparent rise in the viscosity of the normal component at low temperatures discovered by Andronikashvili (1948) may really be caused by the superfluid component.

At small pressure heads or in capillaries of bore less than 5×10^{-3} cm., v_3 becomes small and v_2 predominates. v_2 varies very slowly with the pressure gradient and seems practically independent of the diameter of the capillary. It is, in fact, the velocity represented by that part of figure 3 for which d is greater than about 5×10^{-4} cm. and its order of magnitude is 3 cm. sec.⁻¹ when $\text{grad } p$ is 0.75 dyne cm.⁻³. A velocity of this type is well explained by the hypothesis of Gorter & Mellink (1949) that there is a mutual friction between the two components varying as the n th power of their relative velocity, for, since the normal component must be at rest in capillaries of small bore, the pressure gradient necessary to move the superfluid component against this mutual friction would be determined only by the velocity and not by the diameter of the capillary. The velocity of the superfluid component would, in fact, vary as the $1/n$ th power of the pressure gradient. Gorter & Mellink take n to be approximately 3, but Allen & Misener's experiments with channel widths of 3.2×10^{-3} and 3.9×10^{-4} cm. show a variation of velocity as the $\frac{1}{8}$ th power of the pressure gradient. It is possible, however, that for these two capillaries the measurements were unduly influenced by film flow or thermal effects.

In the narrowest channels and in the film v_1 is the most important component of the velocity, but v_2 is by no means negligible and has the right order of magnitude to explain the variation with pressure head of the film results shown in the filling curve of figure 4. Strictly speaking, then, the velocity previously called v_c may not be v_1 but $v_1 + v_2$, in which case v_1 is difficult to measure exactly except, perhaps, by the method suggested earlier of measuring the amplitude of the oscillations. If it is permissible to assume that v_2 is strictly independent of the channel width, the variation of v_1 with d has the form shown by the dotted curve in figure 3, and its rapid increase as d decreases is one of its most striking features. The variation of $\sigma_1 = v_1 d$ with d is more uncertain. It is possible that σ_1 is constant and equal to $\alpha h/m$ as suggested by Bijl *et al.* (1941), and that the variation of σ_c shown in figure 2 is entirely due to v_2 . However, this would require v_2 to be about 15 cm. sec.⁻¹ in the film, which is five times greater than its value in wider capillaries, and so the question is best left open until a more comprehensive investigation can be made over a wider range of d .

The small damping of the film oscillations proves that v_1 is subject to very little frictional retardation and, if this is taken to be its main characteristic, the absence of observable oscillations in the wider capillaries means that v_1 is then very small, as would be expected from the form of the dotted curve in figure 3. Obviously, no experiment can prove that the frictional forces opposing v_1 are accurately zero, but merely that they are much smaller than would be predicted by some theory under consideration. It can be said, for example, that the theory of Gorter & Mellink, as developed by them to explain thermal conduction in channels of intermediate width, would predict a damping of the film oscillations too large by a factor of the order of 10^5 . Actually they do not claim that their theory is applicable to the narrowest channels and ascribe this to mean free path effects, but the discrepancy just mentioned is so large that it suggests that their theory is fundamentally wrong when applied to the film and is a theory of v_2 but not of v_1 .

Finally, it is important to give consideration to Allen & Misener's observations on the flow of helium II through tightly packed jeweller's rouge in which the interstices were probably of dimensions of the order of 10^{-6} cm., which is comparable with the thickness of the film. They found that, although the velocity was probably no greater than it is in the film, the flow varied as the square root of the pressure head and may therefore have been turbulent. It is difficult to reconcile this fact with any theory which requires curl \mathbf{v} to be zero during the flow of the superfluid component. It is conceivable that the rouge contracted away from the wall of the glass tube into which it was packed and that the flow occurred through the gap thus produced, but the point is obviously of sufficient importance to justify further experimental investigation.

I should like to express my thanks to Professor J. F. Allen for suggesting this problem to me and for many helpful discussions. My thanks are also due to Mr D. V. Osborne who assisted with the measurements. During the early stages of the research I received financial assistance from the Ministry of Education and was enabled to complete the work by the award of an I.C.I. Fellowship and a Fellowship at Trinity College, Cambridge.

REFERENCES

- Allen, J. F. & Misener, A. D. 1939 *Proc. Roy. Soc. A*, **172**, 467.
 Andronikashvili, E. 1948 *J.E.T.P.*, U.S.S.R. **18**, 429.
 Atkins, K. R. 1948 *Nature*, **161**, 925.
 Atkins, K. R. 1950 *Proc. Roy. Soc. A*, **203**, 119.
 van den Berg, G. J. & de Haas, W. J. 1949 *Rev. Mod. Phys.* **21**, 524.
 Bijl, A., de Boer, J. & Michels, A. 1941 *Physica*, **8**, 655.
 Bowers, R. & Mendelssohn, K. 1949 *Nature*, **163**, 870.
 Daunt, J. G. & Mendelssohn, K. 1939 *Proc. Roy. Soc. A*, **170**, 423.
 Daunt, J. G. & Mendelssohn, K. 1946 *Nature*, **157**, 829.
 Gorter, C. J. & Mellink, J. H. 1949 *Physica*, **15**, 285.
 Kapitza, P. L. 1941 *J. Phys. U.S.S.R.* **4**, 181.
 Kapitza, P. L. 1942 *J. Phys. U.S.S.R.* **5**, 59.
 Knudsen, M. 1934 *The kinetic theory of gases*, p. 13. London.
 Landau, L. 1941 *J. Phys. U.S.S.R.* **5**, 71.
 Tisza, L. 1938 *Nature*, **141**, 913.
 Tisza, L. 1947 *Phys. Rev.* **72**, 838.

Stability of viscous flow over concave cylindrical surfaces

BY D. MEKSYN

*Mathematics Department, Imperial College of Science and Technology, London**(Communicated by G. Temple, F.R.S.—Received 20 April 1950).*

The stability for three-dimensional disturbances of viscous flow over concave cylindrical surfaces is investigated by the method of asymptotic solutions

The results confirm Görtler's conclusions that these disturbances can only take place on concave surfaces, and that in the critical condition that the parameter $R_\delta(\delta/r)^{\frac{1}{2}}$ remains constant, where R_δ is Reynolds number and δ/r is the ratio of the thickness of the boundary layer to the radius of curvature.

Additionally, it is found that there are many steady states, and the values of $R_\delta(\delta/r)^{\frac{1}{2}}$ are evaluated for the first two states.

For a fixed $R_\delta(\delta/r)^{\frac{1}{2}}$ there is a finite number of steady states.

1. INTRODUCTION

The flow over concave surfaces was theoretically discussed by Görtler (1940) who showed that it may become unstable under three-dimensional disturbances similar to those found by Taylor (1923) in the case of the rotating cylinder.

Görtler integrated the equations of motion by transforming them into two simultaneous integral equations, and then solving them approximately by taking a few points on the velocity curve; and, thus, reducing the integral equations to a system of several simultaneous linear equations.

This method of solution could give, however, only a rough account of the phenomenon.

Several years ago the present writer (Meksyn 1946) discussed the stability of viscous flow between rotating cylinders by the method of asymptotic expansions;

and showed that this method leads to a comparatively simple general solution which completely agreed with Taylor's theoretical and experimental results.

The aim of the present paper is to apply the same method of solution to the problem of stability of viscous flow over concave cylindrical surfaces.

2. EQUATIONS OF MOTION

For the sake of convenience, the derivation of the equations of disturbed motion is given here, following closely Görtler's (1940) treatment and notations.

The main flow is directed along the axis of x , where x is the length of the arc of the surface, which is perpendicular to the axis of the cylinder; y (≥ 0) is normal to the surface, and z is along the axis, δ is the thickness of the boundary layer and $r = \text{const.}$ is the radius of curvature, which is taken as positive for the concave side of the surface.

The first equation of Navier-Stokes is

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{r}{r-y} u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - \frac{uv}{r-y} + w \frac{\partial u}{\partial z} = - \frac{r}{r-y} \frac{1}{\rho} \frac{\partial p}{\partial x} \\ + \nu \left\{ \frac{r^2}{(r-y)^2} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{r-y} \frac{\partial u}{\partial y} - \frac{2r}{(r-y)^2} \frac{\partial v}{\partial x} - \frac{u}{(r-y)^2} \right\}. \end{aligned} \quad (2.1)$$

Assuming that δ/r is very small, and retaining only terms of the order δ/r , we obtain for the equations of motion and continuity

$$\left. \begin{aligned} \frac{\partial u}{\partial t} + v \left(\frac{\partial u}{\partial y} - \frac{u}{r} \right) + w \frac{\partial u}{\partial z} &= \nu \left\{ \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{r} \frac{\partial u}{\partial y} \right\}, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial y} + \frac{u^2}{r} + w \frac{\partial v}{\partial z} &= - \frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \left\{ \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} - \frac{1}{r} \frac{\partial v}{\partial y} \right\}, \\ \frac{\partial w}{\partial t} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} &= - \frac{1}{\rho} \frac{\partial p}{\partial z} + \nu \left\{ \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} - \frac{1}{r} \frac{\partial w}{\partial y} \right\}, \\ \frac{\partial v}{\partial y} - \frac{v}{r} + \frac{\partial w}{\partial z} &= 0, \end{aligned} \right\} \quad (2.2)$$

where u , v , w are the velocities along x , y , z respectively; the undisturbed flow $u = u_0(y)$, $v = 0$, $w = 0$, $p = p_0$ satisfies the above equations.

Bearing in mind that δ/z is very small, we can introduce in (2.2) the usual boundary-layer approximations; namely, since $\partial u/\partial y \sim u/\delta$, $\partial^2 u/\partial y^2 \sim \partial u/\delta \partial y$, we can disregard u/r in comparison with $\partial u/\partial y$ and $\partial u/r \partial y$ against $\partial^2 u/\partial y^2$, the influence of the surface curvature appears mainly through the term u^2/r in the second equation.

Assume now that

$$\left. \begin{aligned} u &= u_0(y) + u_1(y) \cos \alpha z e^{\beta t}, \\ v &= v_1(y) \cos \alpha z e^{\beta t}, \\ w &= w_1(y) \sin \alpha z e^{\beta t}, \\ p &= p_0(y) + p_1(y) \cos \alpha z e^{\beta t}, \end{aligned} \right\} \quad (2.3)$$

where u_1 , v_1 , w_1 are small quantities whose squares and products can be disregarded.

Introducing (2.3) in (2.2) and, making the boundary-layer simplifications indicated above, we find

$$\left. \begin{aligned} \beta u_1 + v_1 \frac{du_0}{dy} &= \nu \left\{ \frac{d^2 u_1}{dy^2} - \alpha^2 u_1 \right\}, \\ \beta v_1 + u_1 \frac{2u_0}{r} + \frac{1}{\rho} \frac{\partial p_1}{\partial y} &= \nu \left\{ \frac{d^2 v_1}{dy^2} - \alpha^2 v_1 \right\}, \\ \beta w_1 - \frac{\alpha}{\rho} p_1 &= \nu \left\{ \frac{d^2 w_1}{dy^2} - \alpha^2 w_1 \right\}, \\ w_1 &= -\frac{1}{\alpha} \frac{dv_1}{dy}. \end{aligned} \right\} \quad (2.4)$$

Eliminating p_1 between the second and third equations, and substituting the value of w_1 , we obtain

$$\left. \begin{aligned} \nu \frac{d^2 u_1}{dy^2} - (\beta + \nu \alpha^2) u_1 &= v_1 \frac{du_0}{dy}, \\ \nu \frac{d^4 v_1}{dy^4} - (\beta + 2\nu \alpha^2) \frac{d^2 v_1}{dy^2} + \alpha^2 (\beta + \nu \alpha^2) v_1 &= -\frac{2\alpha^2 u_0}{r} u_1. \end{aligned} \right\} \quad (2.5)$$

Transforming the variables, and introducing various abbreviations, let

$$\left. \begin{aligned} \eta &= y/\delta, \quad U = u_0/U_0, \quad \sigma = \alpha\delta, \quad \tau^2 = \alpha^2 \delta^2 + \frac{\beta \delta^2}{\nu}, \\ \mu &= 2 \left(\frac{U_0 \delta}{\nu} \right)^2 \frac{\delta}{r}, \quad u' = \left(\frac{U_0 \delta}{\nu} \right)^{-1} u_1, \quad v' = v_1. \end{aligned} \right\} \quad (2.6)$$

Dropping the primes we obtain

$$\left. \begin{aligned} Lu &= \frac{dU}{d\eta} v, \\ L_0 Lv &= -\sigma^2 \mu U u, \end{aligned} \right\} \quad (2.7)$$

where

$$\left. \begin{aligned} L &= \frac{d^2}{d\eta^2} - \tau^2, \\ L_0 &= \frac{d^2}{d\eta^2} - \sigma^2, \end{aligned} \right\} \quad (2.8)$$

and U is the velocity of the mean flow.

The boundary conditions on the surface of the solid body are

$$u_1(0) = v_1(0) = w_1(0) = 0,$$

or, from (2.4),

$$u_1(0) = v_1(0) = v'(0) = 0.$$

Making use of (2.7) we finally obtain for the boundary conditions on the surface of the solid body

$$u(0) = 0, \quad \frac{d^2 u}{d\eta^2} = 0, \quad \frac{d^3 u}{d\eta^3} - \tau^2 \frac{du}{d\eta} = 0 \quad (2.9)$$

at $\eta = 0$.

3. ASYMPTOTIC INTEGRALS

The flow along concave surfaces is similar to the flow in Taylor's (1923) case of rotating cylinders, when the outer cylinder is at rest, and the inner is rotating.

The method of integration was explained in detail in the previous papers (Meksyn 1946), so that only few explanations will be given here.

Eliminating u between (2.7) we obtain

$$\begin{aligned} \frac{1}{U} \frac{d^6 v}{d\eta^6} - \frac{2U'}{U} \frac{d^5 v}{d\eta^5} - \left[\frac{\sigma^2 + 2\tau^2}{U} + \frac{U''}{U'^2} - \frac{2U'^2}{U^3} \right] \frac{d^4 v}{d\eta^4} \\ + \frac{2(\sigma^2 + \tau^2)}{U^2} U' \frac{d^3 v}{d\eta^3} + \left[\frac{(\sigma^2 + \tau^2) U''}{U^2} - \frac{2(\sigma^2 + \tau^2) U'^2}{U^2} + \frac{\tau^2(2\sigma^2 + \tau^2)}{U} \right] \frac{d^2 v}{d\eta^2} \\ - \frac{2\sigma^2 \tau^2}{U^2} U' \frac{dv}{d\eta} - \left[\frac{\sigma^2 \tau^2}{U^2} U'' - \frac{2\sigma^2 \tau^2}{U^3} U'^2 + \frac{\sigma^2 \tau^4}{U} - \sigma^2 \mu U' \right] v = 0, \end{aligned} \quad (3.1)$$

where primes denote differentiation with respect to η .

Let

$$v = C(\eta) e^{\xi(\eta)}, \quad (3.2)$$

where $C(\eta)$ is a slowly varying function of η , and ξ contains the large parameters μ and σ .

Differentiating (3.2) with respect to η , substituting in (3.1) and retaining only the terms of the two highest orders of magnitude, we obtain from the terms of the first order

$$(\xi'^2 - \tau^2)^2 (\xi'^2 - \sigma^2) + \sigma^2 \mu U U' = 0. \quad (3.3)$$

The terms of the next order of magnitude lead to

$$\begin{aligned} C(\eta) \left[\frac{15}{U} \xi'^4 \xi'' - \frac{2U'}{U} \xi'^5 - \frac{(\sigma^2 + 2\tau^2)}{U} 6\xi'^2 \xi'' \right. \\ \left. + \frac{2(\sigma^2 + \tau^2)}{U^2} U' \xi'^3 + \frac{\tau^2(2\sigma^2 + \tau^2)}{U} \xi'' - \frac{2\sigma^2 \tau^2}{U^2} U' \xi' \right] \\ + C'(\eta) \left[\frac{6}{U} \xi'^5 - \frac{(\sigma^2 + 2\tau^2)}{U} 4\xi'^4 + \frac{\tau^2(2\sigma^2 + \tau^2)}{U} 2\xi' \right] = 0. \end{aligned} \quad (3.4)$$

$$\text{Consider the case} \quad \beta = 0, \quad \text{i.e. } \tau^2 = \sigma^2, \quad (3.5)$$

$$\text{and put for shortness} \quad f(\eta) = 6\xi'(\xi'^2 - \sigma^2)^2. \quad (3.6)$$

Then we can transform (3.4) into

$$C(\eta) \left\{ \frac{1}{U} \frac{df}{d\eta} - \frac{2}{3} \frac{U'}{U^2} f \right\} + \frac{2C'}{U} f = 0, \quad (3.7)$$

$$\text{whence, on integrating,} \quad C(\eta) = \frac{A U^{\frac{1}{2}}}{\xi'^{\frac{1}{2}} (\xi'^2 - \sigma^2)}, \quad (3.8)$$

where A is a constant of integration.

To find u , let

$$u = D(\eta) e^{\xi}, \quad (3.9)$$

whence from (2.7) and (3.8)

$$D(\eta) = \frac{U' C(\eta)}{\xi'^2 - \sigma^2} = \frac{A U^{\frac{1}{2}} U'}{\xi'^{\frac{1}{2}} (\xi'^2 - \sigma^2)^2}. \quad (3.10)$$

To complete the integration we have to solve (3.3) which is done for the case when $\beta = 0$, then

$$\xi'^2 = \sigma^2 \left\{ 1 + \omega \left(-\frac{\mu U U'}{\sigma^4} \right)^{\frac{1}{2}} \right\}, \quad (3.11)$$

where $\omega^3 = 1$.

To proceed with the integration we have to specify the velocity profile $U(\eta)$. It was shown by Liepmann (1945) that the velocity profile in the present case follows rather closely Blasius's flat-plate profile.

According to Görtler the final results depend comparatively little on the precise expression of the velocity profile; this also follows from the asymptotic integrals, since they depend on the integral properties of the velocity curve.

To simplify computations we assume accordingly for U a profile consisting of two straight lines, namely,

$$\left. \begin{aligned} U &= \eta & \text{for } 0 \leq \eta \leq 1, \\ U &= 1 & \text{for } \eta \geq 1, \end{aligned} \right\} \quad (3.12)$$

a case treated also by Görtler.

The asymptotic integrals are obviously not valid for the second region.

We shall now closely follow the procedure, and as far as possible the notation of the paper dealing with the case of rotating cylinders (Meksyn 1946a).

Transferring the origin of co-ordinates at $\eta = \frac{1}{2}$, put

$$\left. \begin{aligned} \eta &= \frac{1}{2} + \eta_1, \\ \xi'^2 &= \sigma^2 \left\{ 1 + \left(-\frac{\mu}{2\sigma^4} \right)^{\frac{1}{2}} (1 + 2\eta_1)^{\frac{1}{2}} \right\} \end{aligned} \right\} \quad (3.13)$$

whence

In the final result there appear terms like

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \xi' d\eta_1,$$

and the term of the first degree in ξ' gives no contribution; whence we obtain approximately, as in the case of rotating cylinders,

$$\left. \begin{aligned} \xi &= \pm \sigma (1 + \omega h)^{\frac{1}{2}} \eta_1, \\ -\frac{I^6}{2\sigma^4} &= h^3, \\ -\frac{1}{2} &\leq \eta_1 \leq \frac{1}{2}. \end{aligned} \right\} \quad (3.14)$$

where

Now introduce the following notations, let

$$\left. \begin{aligned} \omega_1 &= e^{\frac{2\pi i}{3}}, & \omega_2 &= e^{-\frac{2\pi i}{3}}, \\ a &= (1 + \omega_1 h)^{\frac{1}{2}}, & \bar{a} &= (1 + \omega_2 h)^{\frac{1}{2}}, \\ a_0 &= (1 + h)^{\frac{1}{2}}, \end{aligned} \right\} \quad (3.15)$$

where \bar{a} can be taken as a complex conjugate of a ; whence

$$\left. \begin{aligned} \xi'_1 &= \sigma a, & \xi'_2 &= \sigma \bar{a}, & \xi'_3 &= -\sigma a, \\ \xi'_4 &= -\sigma \bar{a}, & \xi'_5 &= \sigma a_0, & \xi'_6 &= -\sigma a_0. \end{aligned} \right\} \quad (3.16)$$

In the determinant which follows from the boundary conditions, the factors of the exponentials in u and v (3.8), (3.10) appear symmetrically, and drop out from the final result; we can therefore put, as far as the *characteristic equation* is concerned,

$$\left. \begin{aligned} u_1 &= \exp(a\sigma\eta), & u_2 &= \exp(\bar{a}\sigma\eta), \\ u_3 &= \exp(-a\sigma\eta), & u_4 &= \exp(-\bar{a}\sigma\eta), \\ u_5 &= \exp(a_0\sigma\eta), & u_6 &= \exp(-a_0\sigma\eta), \end{aligned} \right\} \quad (3.17)$$

where the subscript at η is dropped.

4. INTEGRALS WITHIN THE REGION $U = \text{const.}$

The asymptotic integrals are not valid at the edge of the boundary layer since U' vanishes, and the large parameter μ becomes ineffective. We have, therefore, to integrate the equations for $U = \text{const.}$, and link up the solution with the asymptotic integrals; this provides us with the required three boundary conditions.

At $\eta = 1$,

$$U = 1,$$

and the equations (2.7) become

$$\left. \begin{aligned} \frac{d^4 v}{d\eta^4} - 2\sigma^2 \frac{d^2 v}{d\eta^2} + \sigma^4 v &= -\sigma^2 \mu u, \\ \frac{d^2 u}{d\eta^2} - \sigma^2 u &= 0. \end{aligned} \right\} \quad (4.1)$$

The integrals which tend to zero for η positive and large are

$$\left. \begin{aligned} u_\delta &= A e^{-\sigma\eta}, \\ v_\delta &= (B + C\eta) e^{-\sigma\eta} - \frac{A\mu}{8\eta^2} e^{-\sigma\eta}, \end{aligned} \right\} \quad (4.2)$$

where A, B, C are constants, and the subscript δ denotes that the integrals are taken at the edge of the boundary layer.

The elimination of these constants provides us with three additional boundary conditions.

At the edge of the boundary layer, where the solutions are linked up, the velocities u, v, w have to be equal for both solutions, i.e. from (2.4)

$$u, \quad v, \quad v' \quad (4.3)$$

must be equal for both solutions.

From the first (4.2) we find

$$\sigma u_\delta + u'_\delta = 0, \quad (4.4)$$

and differentiating v_δ with respect to η , we obtain

$$\left. \begin{aligned} \sigma^2 v_\delta + 2\sigma v'_\delta + v''_\delta + \frac{\mu u_\delta}{4} &= 0, \\ \sigma^2 v'_\delta + 2\sigma v''_\delta + v'''_\delta + \frac{\mu u'_\delta}{4} &= 0. \end{aligned} \right\} \quad (4.5)$$

5. THE CHARACTERISTIC EQUATION

Let the general solution of u be

$$u = \sum_{n=1}^b A_n u_n, \quad (5.1)$$

where A_n are constants and u_n the particular integrals (3.17). Introduce the notations (for $\beta = 0$)

$$\left. \begin{aligned} \epsilon_1 = \epsilon_3 = a^2 \sigma^2, \quad \epsilon_2 = \epsilon_4 = \bar{a}^2 \sigma^2, \quad \epsilon_5 = \epsilon_6 = a_0^2 \sigma^2, \\ \eta_1 = -\eta_3 = a \sigma^3 (a^2 - 1), \quad \eta_2 = -\eta_4 = \bar{a} \sigma^3 (\bar{a}^2 - 1), \\ \eta_5 = -\eta_6 = a_0 \sigma^3 (a_0^2 - 1), \end{aligned} \right\} \quad (5.2)$$

the boundary conditions (2.9) on the solid body lead to

$$\sum_{n=1}^6 A_n u_n^{(1)} = 0, \quad \sum_{n=1}^6 A_n \epsilon_n u_n^{(1)} = 0, \quad \sum_{n=1}^6 A_n \eta_n u_n^{(1)} = 0, \quad (5.3)$$

where $u_n^{(1)}$ denotes that it is taken on the boundary of the solid body, i.e. at $\eta = -\frac{1}{2}$.

From (2.7), (3.17), (4.4) and (4.5) we obtain the boundary conditions on the edge of the boundary layer

$$\sum_{n=1}^6 A_n (1 + a_n) u_n^{(2)} = 0, \quad \sum_{n=1}^6 A_n \gamma_n u_n^{(2)} = 0, \quad \sum_{n=1}^6 A_n a_n \gamma_n u_n^{(2)} = 0, \quad (5.4)$$

where

$$\gamma_n = (1 + a_n)^2 (a_n^2 - 1) + \frac{\mu U'}{4\sigma^2}.$$

In the present case $U' = 1$, and $u_n^{(2)}$ denotes that it is taken at $\eta = \frac{1}{2}$.

Eliminating the constants A_n from the six boundary equations we obtain the characteristic equation

$$\begin{vmatrix} u_1^{(1)} & u_2^{(1)} & \dots & u_6^{(1)} \\ \epsilon_1 u_1^{(1)} & \epsilon_2 u_2^{(1)} & \dots & \epsilon_6 u_6^{(1)} \\ \eta_1 u_1^{(1)} & \eta_2 u_2^{(1)} & \dots & \eta_6 u_6^{(1)} \\ (1 + a_1) u_1^{(2)} & (1 + a_2) u_2^{(2)} & \dots & (1 + a_6) u_6^{(2)} \\ \gamma_1 u_1^{(2)} & \gamma_2 u_2^{(2)} & \dots & \gamma_6 u_6^{(2)} \\ a_1 \gamma_1 u_1^{(2)} & a_2 \gamma_2 u_2^{(2)} & \dots & a_6 \gamma_6 u_6^{(2)} \end{vmatrix} = 0. \quad (5.5)$$

In the above expressions only the terms of the highest order are retained.

6. EVALUATION OF THE DETERMINANT

In evaluating the determinant we must bear in mind the order of magnitude of the terms; the parameter a_0 is purely imaginary, but a and \bar{a} are complex quantities, the real part of which can be assumed to be positive.

In the determinant there will be terms of different orders of magnitude; since $\exp(a\sigma)$ is large only terms of the highest order need be retained; they are proportional to

$$u_1^{(2)} u_2^{(2)} u_3^{(1)} u_4^{(1)} u_5^{(2)} u_6^{(1)} \sim \exp\{(a + \bar{a})\sigma + a_0\sigma\},$$

and to the complex conjugate terms, this considerably reduces the computations.

The final result is

$$P e^{i\xi} + Q e^{-i\xi} = 0, \quad (6.1)$$

where

$$\left. \begin{aligned} P &= [-(\epsilon_2 - \epsilon_1)\eta_5 + (\epsilon_5 - \epsilon_1)\eta_2 - (\epsilon_5 - \epsilon_2)\eta_1][(1 + \bar{a})\gamma_1\gamma_5(a_0 - a) \\ &\quad - (1 + a)\gamma_2\gamma_5(a_0 - \bar{a})] + (1 + a_0)\gamma_1\gamma_2(\bar{a} - a)[-\eta_2(\epsilon_5 - \epsilon_1) + \eta_1(\epsilon_5 - \epsilon_2)], \\ \text{and} \\ Q &= [(\epsilon_2 - \epsilon_1)\eta_5 + (\epsilon_5 - \epsilon_1)\eta_2 - (\epsilon_5 - \epsilon_2)\eta_1][(1 + a)\gamma_2\gamma_6(\bar{a}_0 - \bar{a}) \\ &\quad - (1 + \bar{a})\gamma_1\gamma_6(\bar{a}_0 - a)] + (1 + \bar{a}_0)\gamma_1\gamma_2(\bar{a} - a)[\eta_2(\epsilon_5 - \epsilon_1) - \eta_1(\epsilon_5 - \epsilon_2)]. \end{aligned} \right\} \quad (6.2)$$

Dividing the equation (6.1) by $(\epsilon_2 - \epsilon_1)\gamma_1\gamma_2\sigma^3\hbar$, bearing in mind that

$$\frac{\epsilon_5 - \epsilon_1}{\epsilon_2 - \epsilon_1} = -\omega_2, \quad \frac{\epsilon_5 - \epsilon_2}{\epsilon_2 - \epsilon_1} = \omega_1, \quad (6.3)$$

and putting

$$\left. \begin{aligned} (1 + \hbar)^{\frac{1}{2}} &= i |1 + \hbar|^{\frac{1}{2}}, \\ a &= (1 + \omega_1 \hbar)^{\frac{1}{2}} = K e^{i\kappa} = (1 + e^{\frac{2\pi i}{3}} \hbar)^{\frac{1}{2}}, \end{aligned} \right\} \quad (6.4)$$

where $K \cos \kappa > 0$, we obtain, after some lengthy computations,

$$\frac{\tan \sigma |1 + \hbar|^{\frac{1}{2}}}{|1 + \hbar|^{\frac{1}{2}}} = \frac{AD - BKE + 2K^2 \sin \kappa \cos(\kappa - \frac{2}{3}\pi)}{AKE + |1 + \hbar| [BD - 2K^2 \sin \kappa \cos(\kappa - \frac{2}{3}\pi)]}, \quad (6.5)$$

where

$$\left. \begin{aligned} A &= 2(2 + \hbar - \frac{1}{2}\hbar^2)K \cos(\kappa - \frac{2}{3}\pi) - 2|1 + \hbar|, \\ B &= 4K \cos(\kappa - \frac{2}{3}\pi) + (2 + \hbar - \frac{1}{2}\hbar^2), \\ \Delta D &= (1 + 2K \cos \kappa + K^2)[\sin \frac{2}{3}\pi + K \cos(\kappa + \frac{2}{3}\pi)] + \frac{1}{2}\hbar^2 K \sin \kappa, \\ \Delta E &= (1 + 2K \cos \kappa + K^2)[\sin(\kappa + \frac{2}{3}\pi) + K \sin(2\kappa + \frac{2}{3}\pi)] - \frac{1}{2}\hbar^2 \sin \kappa, \\ \Delta &= (1 + 2K \cos \kappa + K^2)^2 - \hbar^2 [\cos \kappa + 2K \cos(\kappa + \frac{2}{3}\pi) + K^2 \cos(2\kappa + \frac{2}{3}\pi) + \frac{1}{4}\hbar^4]. \end{aligned} \right\} \quad (6.6)$$

We have assumed in the above that \hbar is negative, it is easy to see from (6.5) that only under this condition can there be a real solution for σ . It follows, therefore, that such motion can exist only on concave surfaces; a result found by Görtler, and which follows from Taylor's, work on rotating cylinders.

7. SOLUTION OF THE CHARACTERISTIC EQUATION ($\beta = 0$)

We have to solve the equation (6.5) in order to find the critical values of μ and σ corresponding to the minimum Reynolds number when such motion is possible.

Since the right-hand side of (6.5) changes slowly with \hbar , the problem is reduced to finding the minimum value of μ , where

$$\mu \sim \hbar^3 \sigma^4$$

under the condition that $|1+h|^{\frac{1}{2}}\sigma \simeq \text{const.}$,

which leads to $h = -3$.

To solve the equation (6.5) we assume for h a set of values; this enables us to find σ , whence μ is found from

$$\frac{\mu}{2\sigma^4} = -h^3, \quad (7.1)$$

and the critical parameter $R_\delta \sqrt{(\delta/r)}$ from (2.6).

(1) Taking values near $h = -3$ we can find the critical state.

The first two critical states correspond to

$$\left. \begin{aligned} \alpha\delta = 2.4, \quad R_\delta \sqrt{\frac{\delta}{r}} = 54, \quad h = -4.5, \\ \alpha\delta = 4.5, \quad R_\delta \sqrt{\frac{\delta}{r}} = 157, \quad h = -4.0. \end{aligned} \right\} \quad (7.2)$$

If we express the results in terms of momentum thickness θ , where

$$\theta = \int_0^\delta (1-u) u d\eta = \frac{1}{6}\delta,$$

we obtain the corresponding values

$$\left. \begin{aligned} \alpha\theta = 0.400, \quad R_\theta \sqrt{\frac{\theta}{r}} = 3.65, \\ \alpha\theta = 0.75, \quad R_\theta \sqrt{\frac{\theta}{r}} = 10.7. \end{aligned} \right\} \quad (7.3)$$

The results corresponding to the first critical state are of the same order of magnitude as in Taylor's case, where, of course, there are many steady states, although only one solution was given in the case of rotating cylinders.

Görtler (1940) found for the first critical state

$$R_\theta \sqrt{\frac{\theta}{r}} \sim 0.58, \quad \alpha\theta \sim 0.14.$$

The present investigation does not provide any evidence of the existence of this lower state.

Liepmann (1945) investigated experimentally the three-dimensional instability of flow on concave surfaces; according to his results the flow breaks down when

$$R_\theta \sqrt{\frac{\theta}{r}} = 9.0 \quad \text{for } 0.06 \% \text{ turbulence,}$$

and when

$$R_\theta \sqrt{\frac{\theta}{r}} = 6.0 \quad \text{for } 0.3 \% \text{ turbulence.}$$

In figure 1 are given the curves of the first and second steady states.

The upper parts of the curves can only be taken as a very rough approximation, since they correspond to values of h which are very close to unity; for such values, however, the integrals have a critical point in the middle of the flow, and the present

method is not applicable; in this case we have to use a more involved method, as explained in the case of contra-rotating cylinders (Meksyn 1946*b*, p. 492).

(2) As was mentioned above, in order that the equation (6.5) may have a real solution it is necessary that $(1+h)$ should be negative, i.e. the absolute value of h should be larger than unity; whence from (2.6) and (7.1) we obtain

$$\sigma^2 < R_\delta \sqrt{\frac{\delta}{r}},$$

i.e. for a fixed value of R_δ , σ has an upper limit, i.e. the wave-length has a lower limit; this is in striking contrast with the case of vibrations in an inviscid fluid.

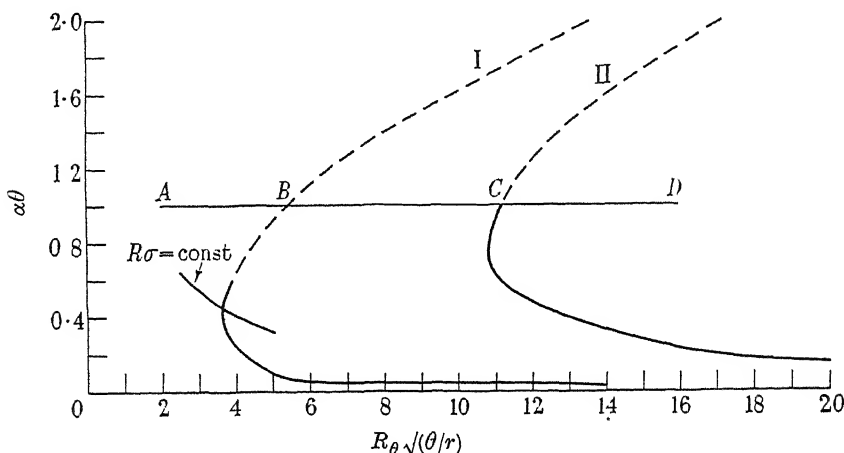


FIGURE 1

(3) Since this kind of instability depends on the integral properties of the velocity curve it should be very little influenced by pressure gradient, as was found experimentally by Liepmann (1945).

This is in great contrast to stability of laminar flow when disturbed by Tollmien's waves, since the latter depend on the differential properties of the velocity curve and are, therefore, very sensitive to anything which changes the form of the velocity curve; as for instance, pressure gradient or suction.

8. APPROXIMATION IN THE SOLUTION

We shall now consider the approximations involved in the above solution.

In the present case, as well as in the analogous case of rotating cylinders, when the external cylinder is at rest, the asymptotic integrals have singularities close to the boundary where ξ' and $\xi'^2 - \sigma^2$ vanish. The solution is not valid close to these points, and we need, therefore, to find the power expansion near these points and link it up with the asymptotic integrals; that is an extremely laborious procedure.

According to Taylor's (1923) measurements, in one case the ratio of the angular velocities of the external and internal cylinders was

$$\mu = \Omega_2/\Omega_1 = 0.160;$$

yet this did not change the critical velocity Ω_1 , as compared with the case when Ω_2 vanishes, and only a comparatively small change occurred in the critical Ω_1 when $\mu = 0.2$.

But in these two cases the integrals have no singularities; we can therefore conclude, on physical grounds, that these singularities have no influence on the final result, they were accordingly eliminated by suitable simplifications of the integrals, as can be seen from the case of rotating cylinders (Meksyn 1948*a*) and from (3.13).

In the present case the position is no less favourable. If we consider, to fix ideas, the first stable state, the critical point is about $1/180$ from the solid boundary, and at distance 0.06 from the boundary the asymptotic integrals are valid.

We may, therefore, safely conclude that the singularities may be eliminated by slightly modifying the mean flow.

The second approximation introduced in the solution of the problem of rotating cylinders, and followed up in the present paper, consists in assuming that ξ' remains virtually constant, having the value corresponding to the middle point of the flow (3.14).

It can be easily shown that, in fact, ξ' changes comparatively little starting, say, from $1/16$ from the boundary to the edge of the boundary layer, and that in the middle point ξ' is equal to the arithmetic mean of its extreme values.

This approximation turned out to be very successful in the case of rotating cylinders, where the boundary conditions are symmetric with respect to both boundaries. In the present case the approximation is, perhaps, less certain because of the dissymmetry of the boundary conditions.

The third approximation was introduced by making use of the linear profile instead of the Blasius's profile.

According to Görtler this has little influence on the final result, and this can also be inferred at a glance from the expressions for the integrals.

In the case of Blasius's profile the integrals have a singularity also at the edge of the boundary layer where U' vanishes; and the profile will require some modification within this region.

It is not difficult to see why all such modifications have little influence on the final result.

The decisive expression is an integral across the boundary layer of what is virtually the sixth root of the velocity, and it is obvious that very slight modifications at the two extremities of the velocity profile, which, on physical grounds, should have no great influence on the motion, are sufficient to get rid of all singularities.

The main aim of this paper is to bring out the general method of integration, and the salient features of this kind of motion; and to this end it is best to consider a case as simple as possible.

9. SOLUTION OF THE CHARACTERISTIC EQUATION, $\beta \neq 0$

We consider now the case when β is not equal to zero (the solution depends on time), but $\beta\delta^2/\nu\sigma^2$ is sufficiently small, so that τ^2 (2.6) differs only slightly from σ^2 ; in that case the main influence of β on the solutions comes from the term $\tan \sigma |1 + h|^{1/2}$ in (6.5).

To integrate the equations of motion we have first to find the roots of the equation (3.3).

$$\left. \begin{aligned} (\xi'^2 - \sigma^2)(\xi'^2 - \tau^2)^2 + \mu U_0'^2 \sigma^2 \eta &= 0, \\ \tau^2 &= \sigma^2 + \frac{\beta \delta^2}{\nu}. \end{aligned} \right\} \quad (9.1)$$

$$\text{Putting} \quad \xi'^2 = \frac{2\tau^2 + \sigma^2}{3} + \frac{\tau^2 - \sigma^2}{3} x, \quad (9.2)$$

$$\left. \begin{aligned} \text{we find} \quad & x^3 - 3x + 2 + H = 0, \\ \text{where} \quad & H = 27p^6(\nu/\beta\delta^2)^3, \\ & p^6 = \mu U_0'^2 \sigma^2 \eta \equiv \frac{1}{2} \mu \sigma^2, \end{aligned} \right\} \quad (9.3)$$

since in the present case $U_0' = 1$, and we consider the solution for the point $\eta = \frac{1}{2}$.

From (9.1), (9.2) and (9.3) we obtain

$$\left. \begin{aligned} \xi'^2 &= \sigma^2[1 + \gamma(x + 2)], \\ \text{where} \quad \gamma &= \frac{\beta \delta^2}{3\nu\sigma^2} \ll 1. \end{aligned} \right\} \quad (9.4)$$

Under these conditions H is very large, whence from (9.3)

$$x \cong -H^{\frac{1}{3}} = -\frac{p^2}{\gamma\sigma^2} = \frac{h}{\gamma}. \quad (9.5)$$

Combining (9.4) and (9.5), and, bearing in mind that γ is small, we obtain

$$\xi' = \sigma(1 + \gamma) \left[1 + \frac{h}{(1 + 2\gamma)} \right]^{\frac{1}{2}}.$$

Since

$$h^3 = -\frac{\mu}{2\sigma^4},$$

$$\left. \begin{aligned} \text{we easily find} \quad & \xi' = \sigma_0(1 + h_0)^{\frac{1}{2}}, \\ \text{where} \quad & \sigma_0 = (1 + \gamma)\sigma, \quad \mu_0 = \frac{\mu}{(1 + \gamma)^2}, \quad h_0^3 = -\frac{\mu_0}{2\sigma_0^4}, \end{aligned} \right\} \quad (9.6)$$

which is identical with the corresponding expression for ξ' , (3.14), in the case when $\beta = 0$, whence, for instance, the solution (7.2) holds good in the present case, i.e.

$$h_0 = -4.5, \quad \alpha_0 \delta = \sigma_0 = 2.4.$$

$$\text{From (9.6)} \quad \sigma = \frac{\sigma_0}{1 + \gamma}, \quad \mu = \mu_0(1 + \gamma)^2, \quad (9.7)$$

where μ_0 and σ_0 correspond to the case $\beta = 0$, finally we get

$$\left. \begin{aligned} R_\delta \sqrt{\frac{\delta}{r}} &= R_{\delta,0} \sqrt{\left(\frac{\delta}{r}\right)} (1 + \gamma), \\ R\sigma &= R_0\sigma_0, \end{aligned} \right\} \quad (9.8)$$

$$\text{whence} \quad \text{if } \beta > 0, \quad \text{then } \sigma < \sigma_0, \quad \text{and } R > R_0, \quad (9.9)$$

i.e. when Reynolds number, or the parameter $R\sqrt{(\delta/r)}$, increases, the motion becomes unstable, and the wave-length increases, the connexion between R and σ being expressed by a hyperbola, the value of β close to the curve of instability is

$$\frac{\beta\delta}{U} = \frac{3\gamma\sigma_0^2}{(1+\gamma)^3R_0} \equiv \frac{3\gamma\sigma_0^2}{R_0}, \quad (9.10)$$

where σ_0 corresponds to the case $\beta = 0$.

If we take a point on the first curve in figure 1 and draw a hyperbola

$$R\sigma = R_0\sigma_0 = \text{const.},$$

the points on this hyperbola inside the curve correspond to larger R and smaller σ , i.e. to unstable states, whereas the points outside the curve correspond to stable states.

If a disturbance starts from a point A , and follows along the line $\sigma = \text{const.}$, it is damped up to B , and destabilized from B , β first increases from B , reaches a maximum value, and is equal to zero at C and so on.

Liepmann (1945) found that there is a continuous change from transition to turbulence due to Görtler's vortices to transition caused by Tollmien's waves.

If we plot in the figures the corresponding curve due to Tollmien's (1929) waves this conclusion can be easily understood.

When r is 'small', Tollmien's curve will move far to the right in the figures, and stability will be dominated by three-dimensional disturbances; by increasing r , Tollmien's curve will move to the left, and the motion is destabilized partly by three-dimensional, and partly by two-dimensional disturbances; at first the former will break up the flow, until by further increase in r Tollmien's curve will move so far to the left, that finally the two-dimensional disturbances will dominate the flow.

It is a pleasure to me to express my thanks to Professor G. Temple, F.R.S., for his very helpful interest in my work, and to the Department of Scientific and Industrial Research for a grant which enabled me to carry out this work.

REFERENCES

- Görtler, H. 1940 *Nachr. Ges. Wiss. Göttingen*, p. 1.
 Liepmann, H. W. 1945 *N.A.C.A. Wartime Report*, W. 87.
 Meksyn, D. 1946a *Proc. Roy. Soc. A*, **187**, 115.
 Meksyn, D. 1946b *Proc. Roy. Soc. A*, **187**, 480.
 Taylor, G. I. 1923 *Phil. Trans. A*, **223**, 289.
 Tollmien, W. 1929 *Nachr. Ges. Wiss. Göttingen*, p. 21.

On the Bose-Einstein condensation

By S. R. DE GROOT,* G. J. HOOYMAN* AND C. A. TEN SELDAM†

(Communicated by E. A. Guggenheim, F.R.S.—Received 17 October 1949.—

Revised 24 April 1950)

The Bose-Einstein condensation of a gas is investigated. Starting from the well-known formulae for Bose statistics, the problem has been generalized to include a variety of potential fields in which the particles of the gas move, and the number w of dimensions has not been restricted to three. The energy levels are taken to be

$$\epsilon_i \equiv \epsilon_{s_1, \dots, s_w} = \text{constant} \frac{\hbar^2}{m} \left\{ \frac{s_1^\alpha - 1}{a_1^2} + \dots + \frac{s_w^\alpha - 1}{a_w^2} \right\} \quad (1 \leq \alpha \leq 2),$$

the quantum numbers being $s_1, \dots, s_w = 1, 2, \dots$, and a_1, \dots, a_w being certain characteristic lengths. (For $\alpha = 2$, the potential field is that of the w -dimensional rectangular box; for $\alpha = 1$, we obtain the w -dimensional harmonic oscillator field.)

A direct rigorous method is used similar to that proposed by Fowler & Jones (1938). It is shown that the number $q = w/\alpha$ determines the appearance of an Einstein transition temperature T_0 . For $q \leq 1$ there is no such point, while for $q > 1$ a transition point exists. For $1 < q \leq 2$, the mean energy $\bar{\epsilon}$ per particle and the specific heat $d\bar{\epsilon}/dT$ are continuous at $T = T_0$. For $q > 2$, the specific heat is discontinuous at $T = T_0$, giving rise to a λ -point.

A well-defined transition point only appears for a very large (theoretically infinite) number N of particles. T_0 is finite only if the quantity $\nu = N/(\alpha_1 \dots \alpha_w)^{2/\alpha}$ is finite. For a rectangular box, ν is equal to the mean density of the gas. If ν tends to zero or infinity as $N \rightarrow \infty$, then T_0 likewise tends to zero or infinity.

In the case $q > 1$, and at temperatures $T < T_0$, there is a finite fraction N_0/N of the particles, given by $N_0/N = 1 - (T/T_0)^q$, in the lowest state. London's formula (1938 *b*) for the three-dimensional box is an example of this equation.

Some further results are also compared with those given by London's continuous spectrum approximation.

PHYSICAL TREATMENT

1. INTRODUCTION

We consider N particles (which are considered to be structureless non-interacting mass-points) distributed over various quantum states in such a way that the number N_i of particles in the state i of energy ϵ_i is given by

$$N_i/(N_i + 1) = \lambda e^{-\epsilon_i/kT} \quad \text{or} \quad N_i = \frac{\lambda e^{-\epsilon_i/kT}}{1 - \lambda e^{-\epsilon_i/kT}}. \quad (1.1)$$

The total number of particles is

$$N = \sum_i N_i = \sum_i \frac{\lambda e^{-\epsilon_i/kT}}{1 - \lambda e^{-\epsilon_i/kT}}. \quad (1.2)$$

The index i is not simply the energy-level number, but denotes the group of quantum numbers which define a micro-state of the particle. Therefore, for two different values of i the energy ϵ_i may sometimes be equal; thus the use of weight factors is unnecessary. To obtain simple results, e.g. a finite λ for $T = 0$, the zero

* Institute for Theoretical Physics, The University, Utrecht, Netherlands.

† Van der Waals Laboratory, The University, Amsterdam, Netherlands.

point of the energy scale is chosen at the lowest level ϵ_0 , which is supposed to be single. The corresponding term N_0 in (1.2) is now

$$N_0 = \lambda/(1 - \lambda), \quad (1.3)$$

which implies $\lambda < 1$.

We call (1.1) the Bose-Einstein distribution. We are not concerned with its foundation or accuracy (for a discussion of these see Schubert 1946; Sommerfeld 1946; Leibfried 1947, Kaempffer 1949; and Wergeland 1944). Our problem is to take the equations (1.2) and (1.3) as defining λ and to examine how the behaviour of λ and thence of the other derived properties as functions of the absolute temperature T and mean density of the gas consisting of these particles depend on the distribution of the quantum states, i.e. on the potential field in which the particles move. Examples of such derived properties are the mean energy of a particle

$$\bar{\epsilon} = \frac{1}{N} \sum_i N_i \epsilon_i = \frac{1}{N} \sum_i \frac{\epsilon_i \lambda e^{-\epsilon_i/kT}}{1 - \lambda e^{-\epsilon_i/kT}}, \quad (1.4)$$

the heat capacity per particle $d\bar{\epsilon}/dT$, (1.5)

and the P - V - T diagram of the gas. These problems are investigated in the case of the rectangular box and of the harmonic oscillator, both in one, two and three dimensions. In the mathematical portion of the paper we investigate more general fields. Many equations in the physical portion of the paper are proved in the mathematical portion. In such cases the references to the corresponding equations of the mathematical portion are given on the left-hand side of the page.

Einstein (1925) was the first to remark that with a perfect gas, consisting of a large number N of particles with mass m , enclosed in a three-dimensional box of volume V , there exists a transition temperature T_0 defined by

$$\frac{N}{V} \left(\frac{h^2}{2\pi m k T_0} \right)^{3/2} = \sum_{j=1}^{\infty} j^{-3/2} = 2.612, \quad (1.6)$$

such that at temperatures below T_0 there appears something like a condensed phase. London (1938*a, b*) calculated the P - V - T diagram of such a gas and showed that at a certain chosen temperature T_0 and volumes smaller than the volume V given by (1.6), the pressure is independent of the volume, as with a real gas in the two-phase region. Therefore, this effect is called the Bose-Einstein condensation. Other investigators, e.g. Uhlenbeck (1927), have raised objections against this nomenclature, reasoning that at the transition point there is not a discontinuity in the energy $\bar{\epsilon}$ (i.e. there is no latent heat) nor even in the specific heat; the discontinuity occurs in $d^2\bar{\epsilon}/dT^2$. Secondly, there is no spatial separation into the two phases as with vapour and liquid, the two phases both fill the whole space as with two mixed gases. Thirdly, as London has remarked, with a Van der Waals gas the $P(V)$ isotherm would not show the above-mentioned horizontal portion. For a perfect gas, however, the name 'condensation' is not altogether inappropriate, and we shall use it throughout our paper.

We shall not approximate to the sums (1.2) and (1.4) by integrals (the continuous spectrum approximation) as is often done (London 1938*b*), but shall investigate

these sums themselves. This has already been carried out by Fowler & Jones (1938) in a somewhat different way for the three-dimensional box, but our aim in the physical portion of the paper is to investigate whether a transition point exists in six special cases, and if so how its character can be described. Some of the results have already been reported in a letter by De Groot & Ten Seldam (1949). One of the principal aims of our investigation is to give a rigorous mathematical basis to the calculations on the Bose-Einstein condensation. Details are summarized in the mathematical portion of the paper.

When $T = 0$ all summands of (1.2) vanish except N_0 , so that $N = N_0$. Hence all the particles are in the lowest state and, because of (1.2),

$$\lambda = N/(N+1). \quad (1.7)$$

In the limiting case $N \rightarrow \infty$,
$$\lambda = 1 \quad (1.8)$$

at $T = 0$. The mean energy $\bar{\epsilon}$ vanishes at $T = 0$.

When T is increased above zero, some particles move to higher levels. This migration from the lowest level increases with increasing T , so that N_0 decreases, and because of (1.3) the same is true of λ . When $T \rightarrow \infty$, all the N_i are equal (all states are equally probable), but they are infinitely small for finite N . This implies that N_0 and λ are zero.

2. THE THREE-DIMENSIONAL BOX

Consider N particles in a rectangular box with sides a_1, a_2 and a_3 . With the chosen energy scale, the energy levels are

$$\epsilon_{s,t,u} = \frac{\hbar^2}{8m} \left(\frac{s^2-1}{a_1^2} + \frac{t^2-1}{a_2^2} + \frac{u^2-1}{a_3^2} \right) \quad (s, t, u = 1, 2, \dots). \quad (2.1)$$

The mean density of the gas
$$n = N/(a_1 a_2 a_3) \quad (2.2)$$

is independent of the temperature.

The mathematical investigation for finite values of N and a_1, a_2, a_3 (§§ A and B) leads to the $\lambda(T)$ curve given in figure 1 and shows that $\lambda, \bar{\epsilon}$ and all their derivatives with respect to T are continuous for every T . There is therefore no transition point in this case.

Now let N and a_1, a_2, a_3 tend to infinity in such a way as to keep the density n constant. The $\lambda(T)$ which results (figure 1) is given by

$$(E.4) \quad \lambda \equiv 1 \quad (T \leq T_0), \quad (2.3)$$

$$(E.2) \quad n \left(\frac{\hbar^2}{2\pi m k T} \right)^{\frac{3}{2}} = 2.612 \left(\frac{T_0}{T} \right)^{\frac{3}{2}} = \sum_{j=1}^{\infty} \lambda^j j^{-\frac{3}{2}} \quad (T \geq T_0), \quad (2.4)$$

in which T_0 is determined by

$$(E.3) \quad n \left(\frac{\hbar^2}{2\pi m k T_0} \right)^{\frac{3}{2}} = \sum_{j=1}^{\infty} j^{-\frac{3}{2}} = 2.612. \quad (2.5)$$

The $\lambda(T)$ curve now consists of two parts that join at the transition point T_0 with a discontinuity in $d^2\lambda/dT^2$. Further, for all temperatures

$$(E.21) \quad \bar{\epsilon} = \frac{3kT}{2n} \left(\frac{2\pi m k T}{\hbar^2} \right)^{\frac{3}{2}} \sum_{j=1}^{\infty} \lambda^j j^{-\frac{5}{2}} = \frac{\frac{3}{2}kT}{2.612} \left(\frac{T}{T_0} \right)^{\frac{3}{2}} \sum_{j=1}^{\infty} \lambda^j j^{-\frac{5}{2}}. \quad (2.6)$$

For $T \leq T_0$, because of (2.3),

$$\bar{\epsilon} = \frac{1.341}{2.612} \frac{3}{2} kT (T/T_0)^{\frac{3}{2}}, \quad (2.7)$$

and hence

$$d\bar{\epsilon}/dT = 1.92k (T/T_0)^{\frac{3}{2}}. \quad (2.8)$$

For $T \geq T_0$ it follows from (2.4) that

$$(E.25) \quad \bar{\epsilon} = \frac{3}{2} kT \sum_{j=1}^{\infty} \lambda^j j^{-\frac{5}{2}} / \sum_{j=1}^{\infty} \lambda^j j^{-\frac{3}{2}}. \quad (2.9)$$

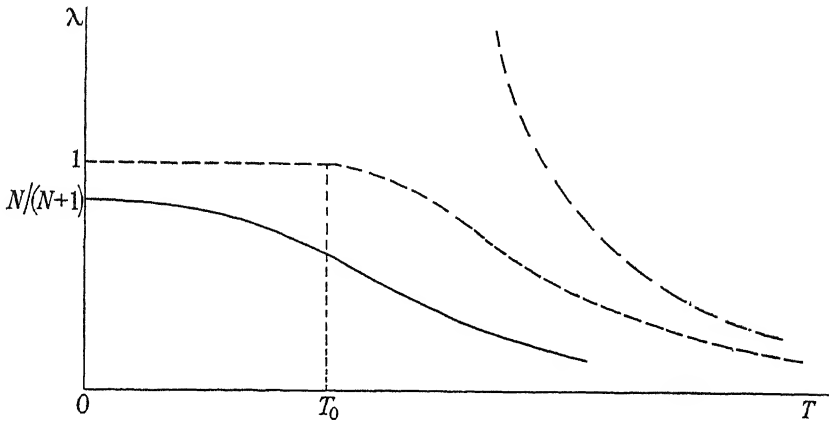


FIGURE 1. $\lambda(T)$ in the case of the three-dimensional box. — finite N and a_1, a_2, a_3 ; --- N infinite, density n finite; - · - curvilinear asymptote; $\lambda = \text{constant } nT^{-\frac{3}{2}}$.

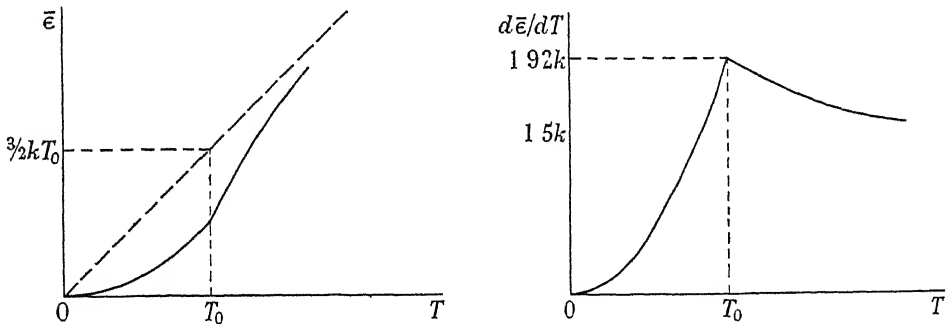


FIGURE 2. The mean energy $\bar{\epsilon}$ and the specific heat $d\bar{\epsilon}/dT$ for the three-dimensional box in the limiting case.

At $T = T_0$, $\bar{\epsilon}$ has a discontinuous second derivative. If $T \rightarrow \infty$, $\lambda \rightarrow 0$, the quotient in (2.9) tends to unity and $\bar{\epsilon} \rightarrow \frac{3}{2} kT$, the classical value (figure 2). For temperatures below T_0 it follows from (1.3) that, in the limiting case, $N_0 \rightarrow \infty$. This happens in such a way that the fraction N_0/N is finite (London 1938*b*):

$$(E.15) \quad n_0/n = N_0/N = 1 - (T/T_0)^{\frac{3}{2}}. \quad (2.10)$$

The density $n_0 = N_0/V$ of the particles in the lowest state is also finite. The number $N - N_0 \equiv N_{\text{eff}}$ of particles not in the lowest state, for which obviously

$$N_{\text{eff}}/N = n_{\text{eff}}/n = (T/T_0)^{\frac{3}{2}} \quad (2.11)$$

may be called the effective particle number. n_{eff} can be defined similarly.

For one mole of particles, $n = \mathcal{N}/\mathcal{V}$, where \mathcal{N} is Avogadro's number and \mathcal{V} the molal volume. Now (2.5) defines T_0 as a function of \mathcal{V} , and (2.3) and (2.4) give the function $\lambda(T, \mathcal{V})$. Finally, the total energy per mole $\mathcal{E} = \mathcal{N}\bar{\epsilon}$ is given by (2.6) as $\mathcal{E}(T, \mathcal{V})$ (according to the energy scale chosen). From this function one can calculate the isotherms as has been done by London (1938*b*). From

$$(\epsilon_i)_{\text{kin}} = \frac{h^2}{8m} \left(\frac{s^2}{a_1^2} + \frac{t^2}{a_2^2} + \frac{u^2}{a_3^2} \right), \quad (2.12)$$

it is clear that, with uniform variation of the volume $\mathcal{V} = a_1 a_2 a_3$, the energy levels are proportional to $\mathcal{V}^{-\frac{2}{3}}$, so that

$$d\epsilon_i/d\mathcal{V} = -\frac{2}{3}\epsilon_i/\mathcal{V}, \quad (2.13)$$

and the pressure is
$$P = -d\mathcal{E}/d\mathcal{V} = \frac{2}{3}\mathcal{E}/\mathcal{V}. \quad (2.14)$$

For $T \geq T_0$ one obtains because of (2.9)

$$P\mathcal{V} = CRT, \quad \text{i.e.} \quad P = nkCT, \quad (2.15)$$

in which C (the quotient of (2.9)) varies from 0.5133 for $T = T_0$ ($\lambda = 1$) to 1 for $T \rightarrow \infty$ ($\lambda \rightarrow 0$). So the gas law is the same as the classical one, except for a factor of the order unity. For $T \leq T_0$, $\lambda = 1$ and

$$P\mathcal{V} = 0.5133RT(T/T_0)^{\frac{1}{3}}, \quad \text{i.e.} \quad P = 0.5133n_{\text{eff}}kT. \quad (2.16)$$

Thus below the transition point a finite fraction n_0/n of the particles is in the lowest state and does not contribute to the energy \mathcal{E} and the pressure P , the really effective particles having a density $n - n_0 = n_{\text{eff}}$. (This explains the name 'effective density' for the quantity n_{eff} .)

According to (2.1) the lowest state $s, t, u = 1$ is taken to have the energy zero. Since the kinetic energy for the lowest level has the value

$$\epsilon_0 = \frac{h^2}{8m} \left(\frac{1}{a_1^2} + \frac{1}{a_2^2} + \frac{1}{a_3^2} \right), \quad (2.17)$$

this means that the potential energy is chosen as $-\epsilon_0$. In our problem s, t and u must each be at least 1. London takes as the lowest state the one with $s, t, u = 0$ and zero kinetic energy. But his boundary conditions are not quite the same as ours. London postulates the periodicity of the wave function in large blocks containing N particles, and normalizes the wave function in such a block. For this boundary problem wave functions exist with some or all of the numbers s, t, u equal to zero. For $s, t, u = 0$ the solution is $\Psi = \text{constant}$. In this case the energy levels are

$$\epsilon_i = \frac{h^2}{8m} \left(\frac{s^2}{a_1^2} + \frac{t^2}{a_2^2} + \frac{u^2}{a_3^2} \right) \quad (s, t, u = 0, 1, 2, \dots), \quad (2.18)$$

where the potential energy is chosen to be zero. If (2.18) is used instead of (2.1), we obtain the same qualitative results. The quantitative results in the limiting case of infinite N with constant n are not affected at all (cf. the remark at the end of §E). In the periodic boundary problem a phenomenon occurs for finite N which in our problem only happens in the case $N \rightarrow \infty$, namely, the particles in the lowest state

(with zero kinetic energy) show a 'condensation in momentum space', and, since their wave functions are constant, they possess a 'peculiar omnipresence in ordinary space'. In our problem, the kinetic energy ϵ_0 of the lowest level tends to zero only when a_1, a_2, a_3 tend to infinity.

3. THE ONE- AND TWO-DIMENSIONAL BOX

The energy of the quantum states is given by the expression (2.1), but now with one or two terms respectively. With finite N , a_1 (and a_2) the general character of $\lambda(T)$ is again given by figure 1 (full line). For $N \rightarrow \infty$ and $a_1(a_2) \rightarrow \infty$ with constant density n (equal to N/a_1 and $N/a_1 a_2$ respectively), however, no transition point now appears.

In the one-dimensional case (figure 3) for all temperatures

$$(E.2) \quad n \sqrt{\{h^2/(2\pi m k T)\}} = \sum_{j=1}^{\infty} \lambda^j j^{-\frac{1}{2}} \quad (3.1)$$

and

$$(E.21) \quad \bar{\epsilon} = \frac{1}{2n} \sqrt{\{(2\pi m k T)/h^2\}} k T \sum_{j=1}^{\infty} \lambda^j j^{-\frac{3}{2}} = \frac{1}{2} k T \sum_j \lambda^j j^{-\frac{3}{2}} / \sum_j \lambda^j j^{-\frac{1}{2}}. \quad (3.2)$$

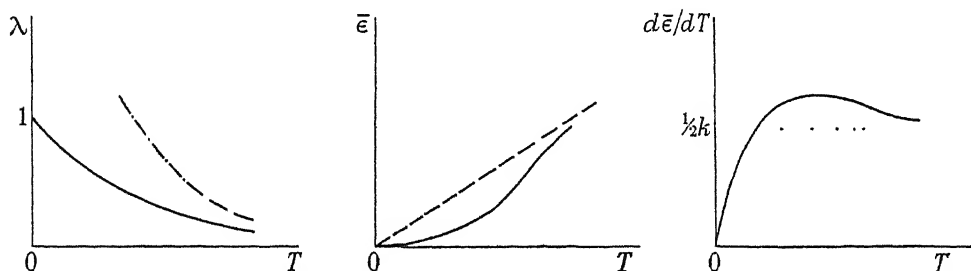


FIGURE 3. λ , $\bar{\epsilon}$ and $d\bar{\epsilon}/dT$ in the case of the one-dimensional box with infinite N and finite density n . - - - asymptote, $\lambda = \text{constant } nT^{-1}$, - - - asymptote, $\bar{\epsilon} = \frac{1}{2}kT$.

In the two-dimensional case (figure 4) for all temperatures

$$(E.2) \quad n h^2 / (2\pi m k T) = \sum_{j=1}^{\infty} \lambda^j j^{-1} = -\ln(1 - \lambda) \quad (3.3)$$

$$\text{and} \quad \bar{\epsilon} = \frac{1}{n} \frac{2\pi m k T}{h^2} k T \sum_{j=1}^{\infty} \lambda^j j^{-2} = k T \sum_j \lambda^j j^{-2} / \sum_j \lambda^j j^{-1}. \quad (3.4)$$

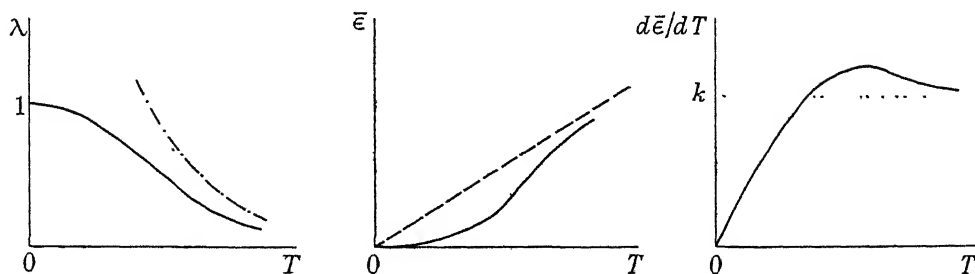


FIGURE 4. λ , $\bar{\epsilon}$ and $d\bar{\epsilon}/dT$ in the case of the two-dimensional box with infinite N and finite density n . - - - asymptote, $\lambda = \text{constant } nT^{-1}$; - - - asymptote, $\bar{\epsilon} = kT$.

4. THE HARMONIC OSCILLATOR, THREE DIMENSIONS

For the sake of simplicity we shall consider in this section only the isotropic case and suppose the potential energy to be

$$V = \frac{1}{2}f(x^2 + y^2 + z^2), \quad (4.1)$$

where x, y, z are the deviations from the origin $(0, 0, 0)$. The energy levels are given by

$$\epsilon_{s,t,u} = \frac{\hbar}{2\pi} \left(\sqrt{\frac{f}{m}} \right) (s+t+u + \frac{3}{2}) \quad (s, t, u = 0, 1, \dots). \quad (4.2)$$

Each particle in the lowest state is, classically speaking, confined to a 'living-space', bounded by the surface

$$\frac{1}{2}f(x^2 + y^2 + z^2) = \frac{3\hbar}{4\pi} \sqrt{\frac{f}{m}}, \quad (4.3)$$

this being a sphere of radius a , where $\frac{1}{2}fa^2$ is given by the right-hand side of (4.3). Now, when the energy scale is chosen as used in the preceding sections, (4.2) may be rewritten in the form

$$\epsilon_{s,t,u} = \frac{3\hbar^2}{4\pi^2 a^2 m} (s+t+u). \quad (4.4)$$

For finite values of N and a , the behaviour of $\lambda(T)$ is again as given in figure 1 (full line). Again λ , $\bar{\epsilon}$ and all their derivatives with respect to T are continuous for every finite T .

(a) *A first limiting case*

For the sake of analogy with the case of the box, we take

$$N = na^3, \quad (4.5)$$

and now let a tend to infinity with constant n .

Let us first consider the physical significance of n . With the problem of the box, for all particles the 'living-space' (quantum mechanical as well as classical) is the box itself and n is obviously the mean density. In the present problem, the classical living-space of a particle depends on its energy, the quantum-mechanical living-space is infinite. The question arises whether n is only a mathematical abstraction in this case. We can only say, for example, that if all particles were in the lowest state and confined to the corresponding classical living-space (the sphere of radius a), the mean density would be $3n/(4\pi)$; or, alternatively, if we consider many systems of particles, each system having the same distribution of its particles over its states, the concentration at, for example, the centre of a system would be proportional to the value of N/a^3 for that system.

The behaviour of $\lambda(T)$ for increasing a is shown in figure 5; for $a = \infty$, $\lambda = 1$ at $T = 0$ and $\lambda = 0$ for all positive T .

The graph of $\lambda(T)$ is explained best with the aid of the graph of λ as a function of the auxiliary quantity $b = aT$. As a and N increase, the $\lambda(b)$ curve deforms in the direction $1 \rightarrow 4$ and meanwhile develops a horizontal left portion and a rather sharp

bend, which tends to a real discontinuity in the derivative $d\lambda/db$. In the limiting case we have

$$(F.2) \quad \left. \begin{aligned} \lambda &\equiv 1 \quad \text{for } b \leq b_0, \\ n(M/b)^3 &= \sum_{j=1}^{\infty} \lambda^j j^{-3} \quad \text{for } b \geq b_0, \end{aligned} \right\} \quad (4.6)$$

where b_0 is determined by

$$n(M/b_0)^3 = \sum_{j=1}^{\infty} j^{-3} = 1.202 \quad (4.7)$$

and

$$M = 3\hbar^2/(4\pi^2 mk). \quad (4.8)$$

The mean energy is

$$(E.26) \quad \bar{e} = 3kT. \quad (4.9)$$

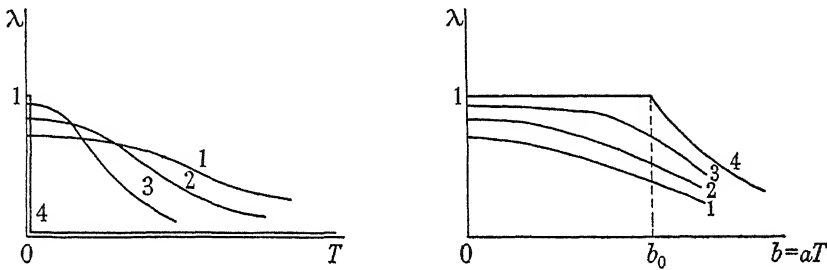


FIGURE 5. λ in the case of the three-dimensional isotropic harmonic oscillator as a function of T and aT respectively for constant n . The curves 1, 2, 3 correspond in that sequence to increasing values of N . Curve 4 corresponds to infinite N .

This result, which can be derived rigorously from the calculations given in the mathematical portion of the paper, can also be found with the 'continuous spectrum' approximation method (§G) as used in the box problem (London 1938*b*).

The absence of quantum-mechanical degeneration is caused by the fact that the transition point, while developing, at the same time tends to zero temperature. The transition temperature T_0 can be used as a characteristic temperature at which the energy is a certain fixed fraction of the classical value (see (E.25)). In view of this property, we may call each finite temperature 'high' in this limiting case.

(b) A second limiting case

If the quantity ν , defined by

$$\nu = n/a^3 = N/a^6, \quad (4.10)$$

is taken to be constant instead of n , we should again find a finite transition temperature in the limiting case $N \rightarrow \infty$, in which

$$(E.4) \quad \left. \begin{aligned} \lambda &\equiv 1 \quad \text{for } T \leq T_0 \\ \nu(M/T)^3 &= \sum_{j=1}^{\infty} \lambda^j j^{-3} \quad \text{for } T \geq T_0, \end{aligned} \right\} \quad (4.11)$$

where M is given by (4.8) and T_0 by

$$(E.3) \quad \nu(M/T_0)^3 = \sum_{j=1}^{\infty} j^{-3} = 1.202. \quad (4.12)$$

The graph of $\lambda(T)$ would show a discontinuity in the derivative $d\lambda/dT$ at $T = T_0$ and have the same shape as the curve 4 of $\lambda(b)$ in the first limiting case (figure 5).

For $\bar{\epsilon}$ we should find

$$(E.21) \quad \bar{\epsilon} = \frac{3kT}{\nu} \left(\frac{T}{M} \right)^3 \sum_{j=1}^{\infty} \lambda^j j^{-4}, \quad (4.13)$$

and so, because of (4.11),

$$\left. \begin{aligned} \bar{\epsilon} &= 1.0823 \frac{3kT}{\nu} \left(\frac{T}{M} \right)^3 & \text{for } T \leq T_0, \\ \bar{\epsilon} &= 3kT \sum \lambda^j j^{-4} / \sum \lambda^j j^{-3} & \text{for } T \geq T_0 \end{aligned} \right\} \quad (4.14)$$

As figure 6 shows, $d\bar{\epsilon}/dT$ would have a discontinuity at $T = T_0$. In this case therefore we should again find a quantum-mechanical degeneration. But in the limiting case the gas would have strange properties, since the 'mean density' n would tend to infinity with $N \rightarrow \infty$.

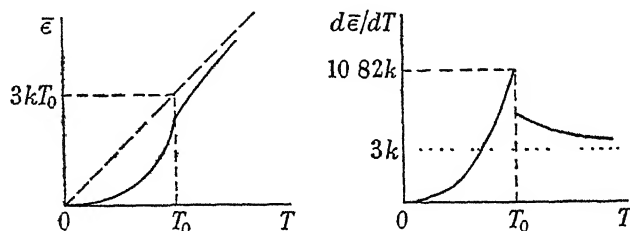


FIGURE 6. $\bar{\epsilon}$ and $d\bar{\epsilon}/dT$ in the case of the three-dimensional isotropic harmonic oscillator for infinite N and constant $\nu = N/a^3$.

We can now state the results in the first limiting case (§4*a*) as follows: With constant n and $N \rightarrow \infty$, ν tends to zero and so does T_0 according to (4.12). [Similar phenomena would be encountered if, in the case of the three-dimensional box, we should keep, for example, N constant instead of n so that n would vanish with increasing a_1, a_2, a_3 . The mean energy $\bar{\epsilon}$ would be equal to $\frac{3}{2}kT$ for all temperatures in that limiting case.]

(c) A third limiting case

We now consider the case of an infinite number N of particles in an isotropic potential field (4.1) with finite a . This problem seems to be more sensible than that in §4*a* because there nothing was left of the oscillator character of the field in the limiting case $a \rightarrow \infty$.

We should then have in the limiting case $N \rightarrow \infty$

$$(\S F) \quad \lambda \equiv 1 \quad \text{for all } T. \quad (4.15)$$

[In terms of §4*b* this result could be stated as follows: With ν tending to infinity, the transition temperature T_0 likewise tends to infinity according to (4.12), so that $\lambda \equiv 1$ for all T .]

The mean energy per particle $\bar{\epsilon}$ would be zero, but the total energy $E = N\bar{\epsilon}$ would have a finite value and only tend to infinity with $a \rightarrow \infty$ or with $T \rightarrow \infty$.

The physical meaning of this is that in this case only a finite number $N - N_0$ of particles are not in the lowest state. This can also be derived from

$$N - N_0 = \sum_{i \neq i_{\text{lowest state}}} \frac{\lambda e^{-\epsilon_i/kT}}{1 - \lambda e^{-\epsilon_i/kT}}, \quad (4.16)$$

where ϵ_i is given by (4.4), since this series converges for finite a even if $\lambda = 1$ (see §F).

[As is shown in §F, the above remarks are equally valid in the case of the box if an infinite number N of particles is confined to a box of finite dimensions.]

5 THE HARMONIC OSCILLATOR IN ONE AND TWO DIMENSIONS

For finite a and N we may argue as in the three-dimensional case, and so λ , $\bar{\epsilon}$ and their derivatives are continuous functions of T .

If, in the one-dimensional case, we first let N and a tend to infinity with constant 'density' $n = N/a$, we find no transition point. Introducing the quantity $b = aT$, we find for all values of b

$$(F.2) \quad nM/b = \sum \lambda^j j^{-1} = -\ln(1 - \lambda), \quad (5.1)$$

where M is given by (4.8). $\bar{\epsilon}$ is equal to the non-degenerate value kT for every positive T , just as in the same three-dimensional limiting case (§4a).

Considering a limiting case similar to that of §4b with constant $\nu = N/a^2$, we should again find a quantum-mechanical degeneration in the energy expression. A transition point does not exist, since for all temperatures

$$(E.2) \quad \nu M/T = \sum \lambda^j j^{-1} = -\ln(1 - \lambda), \quad (5.2)$$

$$(E.21) \quad \bar{\epsilon} = \frac{kT}{\nu} \frac{T}{M} \sum \lambda^j j^{-2} = kT \sum \lambda^j j^{-2} / \sum \lambda^j j^{-1}, \quad (5.3)$$

so that $\bar{\epsilon}$ and $d\bar{\epsilon}/dT$ are exactly the same as in the case of the two-dimensional box with constant n (figure 4).

With the two-dimensional oscillator field, we find, introducing $b = aT$, in the first limiting case ($N \rightarrow \infty$ with constant 'density' $n = N/a^2$) the existence of a transition point for the behaviour of $\lambda(b)$:

$$(F.2) \quad \left. \begin{aligned} \lambda &\equiv 1 & \text{for } b \leq b_0, \\ n(M/b)^2 &= \sum_{j=1}^{\infty} \lambda^j j^{-2} & \text{for } b \geq b_0, \end{aligned} \right\} \quad (5.4)$$

$$\text{with } b_0 \text{ given by} \quad n(M/b_0)^2 = \sum j^{-2} = 1.645. \quad (5.5)$$

The discontinuity appears in $d^2\lambda/db^2$. But again, as in §4a, the transition temperature $T_0 = b_0/a$, while developing, tends to zero. So the quantum-mechanical degeneration is absent in this limiting case (and $\bar{\epsilon} = 2kT$), but it would reappear in the second case of increasing N with constant $\nu = N/a^4$ analogous to §4b.

In the limiting case of §4c ($N \rightarrow \infty$ with finite a), for both the one- and two-dimensional oscillators we should again find (4.15) and the results given there.

6. DISCUSSION

What factor determines the occurrence of a transition point? The answer is that the differences between the various cases are caused by the different behaviour of a certain series, namely, the series

$$R_q(\lambda) \equiv \sum_{j=1}^{\infty} \lambda^j j^{-q}, \quad (6.1)$$

which determines the behaviour of the functions $\lambda(T)$ and $\bar{\epsilon}(T)$ in the limiting case $N \rightarrow \infty$, as will be shown in the mathematical portion (theorems 1 and 3).

In all our cases we have

$$q = w/\alpha, \quad (6.2)$$

where w is the number of dimensions and α is the exponent of the quantum numbers in the expression for the energy ϵ_i (with the box $\alpha = 2$, with the oscillator field $\alpha = 1$).

Now for $q \leq 1$, the series (6.1) diverges for $\lambda = 1$ (see figure 8), and therefore varies monotonically from zero to infinity if λ increases from 0 to 1. This behaviour does not give rise to the occurrence of a transition point.

For $q > 1$, however, (6.1) has a finite value for $\lambda = 1$ and does not exceed this value for $0 \leq \lambda \leq 1$. From this, the existence of a transition temperature T_0 can be deduced. For $q > 2$, $d\lambda/dT$ and the specific heat $d\bar{\epsilon}/dT$ are discontinuous at T_0 . For $1 < q \leq 2$, the discontinuities occur in the second or higher derivatives of λ and $\bar{\epsilon}$.

Our conclusion therefore is that the existence as well as the character of the transition point is chiefly determined by the value of q .

As is shown in the mathematical portion (§A), a general field can be introduced, of which the preceding problems are special cases. London (1939), connecting the λ phenomenon of liquid helium with the Bose-Einstein condensation, suggested as a likely result of an interaction between the atoms a continuous energy spectrum with a density of energy levels given by

$$g(\epsilon) d\epsilon = AV\epsilon^{\sigma-1} d\epsilon, \quad (6.3)$$

where V is the molecular volume, A an adjustable constant and $\sigma > 2$. As a matter of fact, our starting-point also leads to this expression, when we use, as London did, the continuous spectrum approximation. An energy spectrum such as (G.1) then leads to the approximation (G.4), which can be identified with London's expression (6.3), if we take

$$\sigma = q, \quad (6.4)$$

$$AV = (a^2/Mk)^q \{(\alpha^{-1})!\}^w / (q-1)!. \quad (6.5)$$

The reasons for London's choice $\sigma > 2$ will be clear from our mathematical treatment. With energy levels as represented by (A.3), it is only for $q > 2$ that a λ -point (a discontinuity in the specific heat) occurs (cf. the résumé at the end of §E). (London's results for the transition temperature T_0 and the specific heat then coincide with ours, if we identify the number N' , introduced by London, with the total number N of particles.) Differences may arise as to the interpretation of the potential field, leading to the above-mentioned energy states. London considers these states as

'resulting from an interaction between the particles', whereas we speak of 'particles without interaction, moving in an external potential field'. However, as far as the mathematical treatment is concerned, this is only a matter of nomenclature.

MATHEMATICAL TREATMENT

A. GENERALIZATION OF THE PROBLEM

Because $\lambda < 1$, the equations (1.2) and (1.4) can be written in the form

$$N = \sum_{j=1}^{\infty} \lambda^j \sum_i e^{-j\epsilon_{ij}/kT}, \quad (\text{A } 1)$$

$$\bar{\epsilon} = -\frac{kT}{N} \sum_{j=1}^{\infty} \lambda^j \frac{d}{dj} \sum_i e^{-j\epsilon_{ij}/kT}. \quad (\text{A } 2)$$

In order to include the various cases in one single mathematical treatment, we can introduce a general potential field in w dimensions, to which belong the energy levels

$$\frac{\epsilon_{s_1, \dots, s_w}}{kT} = \frac{M}{T} \sum_{v=1}^w \frac{s_v^{\alpha} - 1}{a_v^2} \quad (s_1, \dots, s_w = 1, 2, \dots), \quad (\text{A } 3)$$

where

$$M = \text{constant } h^2/mk, \quad (\text{A } 4)$$

so that the energy scale has its zero point at the lowest level $\epsilon_{1, \dots, 1}$. The constants a_1, \dots, a_w have the dimensions of a length. For $\alpha = 2$, (A.3) represents the energy levels of a particle enclosed in a box; for $\alpha = 1$ the levels are those of a particle in a harmonic oscillator field. We suppose that

$$1 \leq \alpha \leq 2. \quad (\text{A } 5)$$

With this choice of the potential field, (A.1) transforms into

$$N = \sum_{j=1}^{\infty} \lambda^j \prod_{v=1}^w G_{\alpha}(x_v j), \quad (\text{A } 6)$$

and (A.2) into

$$\bar{\epsilon} = -\frac{kT}{N} \sum_{j=1}^{\infty} \lambda^j \frac{d}{dj} \prod_{v=1}^w G_{\alpha}(x_v j), \quad (\text{A } 7)$$

where

$$x_v = M/a_v^2 T \quad (\text{A } 8)$$

and

$$G_{\alpha}(x) = e^x \sum_{s=1}^{\infty} e^{-xs^{\alpha}} = 1 + e^{-x(2^{\alpha}-1)} + e^{-x(3^{\alpha}-1)} + \dots \quad (\text{A } 9)$$

For positive values of x , $G_{\alpha}(x)$ is a continuous, monotonically decreasing function of x , so that

$$\lim_{x \rightarrow 0} G_{\alpha}(x) = +\infty, \quad (\text{A } 10)$$

$$\lim_{x \rightarrow \infty} G_{\alpha}(x) = 1. \quad (\text{A } 11)$$

Likewise, the derivatives $-G'_{\alpha}(x)$ and $G''_{\alpha}(x)$ are continuous and monotonically decreasing functions of x as $x > 0$, such that

$$\lim_{x \rightarrow 0} -G'_{\alpha}(x) = \lim_{x \rightarrow 0} G''_{\alpha}(x) = +\infty, \quad (\text{A } 12)$$

and, for each fixed number p ,

$$\lim_{x \rightarrow +\infty} -x^p G'_\alpha(x) = \lim_{x \rightarrow +\infty} x^p G''_\alpha(x) = 0. \quad (\text{A.13})$$

The first term of (A.13) can be found with the aid of the following inequalities:

$$0 < -G'_\alpha(x) = \sum_{s=2}^{\infty} (s^\alpha - 1) e^{-x(s^\alpha - 1)} < \sum_{s=2}^{\infty} s^2 e^{-x(s-1)} = t \frac{t^2 - 3t + 4}{(1-t)^3}, \quad (\text{A.14})$$

where $t = e^{-x}$.

B. INVESTIGATION FOR FINITE N AND α_v

If, with fixed T , λ increases from 0 to 1, the right-hand side of (A.6) increases monotonically and continuously from zero to positive infinity. So for each positive value of N , equation (A.6) has exactly one solution $\lambda(T)$ between 0 and 1. Because of (A.11) we have

$$\lambda_0 \equiv \lambda(T=0) = N/(N+1), \quad (\text{B.1})$$

$$\text{and because of (A.10)} \quad \lambda(T=\infty) = 0. \quad (\text{B.2})$$

Since $G_\alpha(x)$ is monotonically decreasing, it follows from (A.6) that, within an arbitrarily chosen finite interval of temperatures

$$0 < T \leq \tau, \quad (\text{B.3})$$

λ is greater than a certain positive λ_{\min} , the solution of (A.6) for $T = \tau$, so that

$$0 < \lambda_{\min} \leq \lambda < \lambda_0. \quad (\text{B.4})$$

Differentiating (A.6) with respect to λ , we find $\partial N / \partial \lambda$, which, like the right-hand part of (A.6), is bounded and continuous in each finite interval of T , with

$$\partial N / \partial \lambda > 0. \quad (\text{B.5})$$

Likewise we can form $\partial N / \partial T$, which is bounded and continuous in the T, λ region given by (B.3) and (B.4), in such a way that its value is positive for positive T and vanishes for $T = 0$. From this it follows that the derivative

$$\frac{d\lambda}{dT} = -\frac{\partial N / \partial T}{\partial N / \partial \lambda} \quad (\text{B.6})$$

exists and is continuous in the given region, is negative for $T > 0$ and vanishes for $T = 0$. (In the same way it can be proved that $d^2\lambda/dT^2$ also is bounded and continuous, and is zero for $T = 0$.) From the above statements we conclude that the graph of $\lambda(T)$ is as given, for example, in figure 1 (full line).

Likewise it can be proved that $\bar{\epsilon}$, $d\bar{\epsilon}/dT$ and $d^2\bar{\epsilon}/dT^2$ are bounded and continuous for finite T and vanish for $T = 0$.

C. THE FUNCTION $G_\alpha(x)$ FOR SMALL VALUES OF x

Though $G_\alpha(x)$ tends to infinity for $x \rightarrow 0$, $x^{1/\alpha} G_\alpha(x)$ has the finite limit $(\alpha^{-1})!$, since

$$\lim_{x \rightarrow 0} x^{1/\alpha} G_\alpha(x) = (\alpha^{-1})!. \quad (\text{C.1})$$

The proof is as follows. Taking $x^{1/\alpha} = \Delta z$, we clearly obtain (see figure 7a)

$$x^{1/\alpha} e^{-x} G'_\alpha(x) = \sum_{s=1}^{\infty} e^{-(s\Delta z)^\alpha} \Delta z \rightarrow \int_0^{\infty} e^{-z^\alpha} dz = (\alpha^{-1})! \quad (\Delta z \rightarrow 0), \quad (\text{C.2})$$

and hence (C.1) follows.

Introducing a function $h_\alpha(x)$ defined by

$$\{(\alpha^{-1})!\}^{-1} x^{1/\alpha} G'_\alpha(x) \equiv h_\alpha(x) + \{(\alpha^{-1})!\}^{-1} x^{1/\alpha}, \quad (\text{C.3})$$

we have

$$\lim_{x \rightarrow 0} h_\alpha(x) = 1, \quad (\text{C.4})$$

and, for all $x > 0$,

$$0 < h_\alpha(x) < 1. \quad (\text{C.5})$$

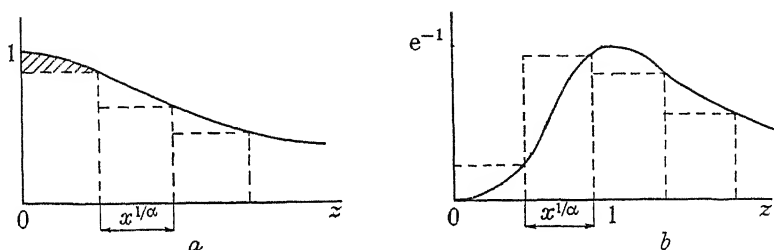


FIGURE 7. The functions. a, e^{-z^α} and b, $z^\alpha e^{-z^\alpha}$.

To prove the second inequality of (C.5) we remark that the difference between $(\alpha^{-1})!$ and the sum of the series in (C.2) is majorant to the area S of the shaded 'triangle' in figure 7a. Now, with $\alpha \geq 1$,

$$\int_{x^{1/\alpha}}^{\infty} e^{-z^\alpha} dz \leq e^{-x} \int_{x^{1/\alpha}}^{\infty} e^{-(z-x^{1/\alpha})^\alpha} dz = (\alpha^{-1})! e^{-x}, \quad (\text{C.6})$$

and so

$$\int_0^{x^{1/\alpha}} e^{-z^\alpha} dz \geq (\alpha^{-1})! (1 - e^{-x}). \quad (\text{C.7})$$

Hence

$$S \geq (\alpha^{-1})! - e^{-x} \{(\alpha^{-1})! + x^{1/\alpha}\}, \quad (\text{C.8})$$

which leads to the second inequality of (C.5).

Similarly, $-G'_\alpha(x)$ tends to infinity like $x^{-1-1/\alpha}$ as $x \rightarrow 0$. To prove this, write

$$-G'_\alpha(x) = -G_\alpha(x) + e^x \sum_{s=1}^{\infty} s^\alpha e^{-xs^\alpha}. \quad (\text{C.9})$$

Then, because of (C.1), and taking $x^{1/\alpha} = \Delta z$ (see figure 7b), we have

$$-\lim_{x \rightarrow 0} x^{1+1/\alpha} G'_\alpha(x) = \lim_{\Delta z \rightarrow 0} \sum_{s=1}^{\infty} (s\Delta z)^\alpha e^{-(s\Delta z)^\alpha} \Delta z = \int_0^{\infty} z^\alpha e^{-z^\alpha} dz = (\alpha^{-1})!/\alpha. \quad (\text{C.10})$$

Introducing

$$K_\alpha(x) = -\alpha \{(\alpha^{-1})!\}^{-1} x^{1+1/\alpha} G'_\alpha(x), \quad (\text{C.11})$$

we have, with fixed positive p , in view of (A.13),

$$K_\alpha(x) \rightarrow 1 \quad (x \rightarrow 0) \quad (\text{C.12})$$

$$x^p K_\alpha(x) \rightarrow 0 \quad (x \rightarrow \infty). \quad (\text{C.13})$$

D. THE FUNCTION $R_q(\lambda)$

As we already remarked in §6, the crucial series, determining the occurrence of a transition point, is

$$R_q(\lambda) \equiv \sum_{j=1}^{\infty} \lambda^j j^{-q}. \quad (\text{D.1})$$

For $0 \leq \lambda < 1$ this series converges for each q , whereas for $\lambda = 1$ it coincides with Riemann's ζ -function, defined for $q > 1$ (see figure 8). The derivative $dR_q/d\lambda$ is given by

$$\frac{dR_q}{d\lambda} = \frac{R_{q-1}(\lambda)}{\lambda} \quad (\text{D.2})$$

and is unity for $\lambda = 0$. For each $q < 1$ and $\lambda \rightarrow 1$, $R_q(\lambda)$ tends to infinity in such a way that

$$(1-\lambda)^{-q+1} R_q(\lambda) \rightarrow (-q)!, \quad (\text{D.3})$$

and consequently, because of (D.2), for $q < 2$,

$$(1-\lambda)^{-q+2} \frac{dR_q}{d\lambda} \rightarrow (-q+1)!. \quad (\text{D.4})$$

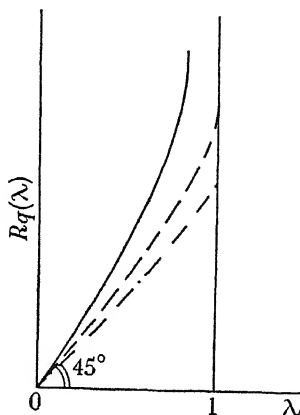


FIGURE 8. The function $R_q(\lambda)$. —, $0 \leq q \leq 1$ (asymptote $\lambda = 1$); ---, $1 < q \leq 2$ (finite value, infinite slope for $\lambda = 1$); - · -, $q > 2$ (finite value and slope for $\lambda = 1$).

To prove this, we take $\lambda = e^{-x}$ and we then have, with $\lambda \rightarrow 1$, $x \rightarrow +0$,

$$(1-\lambda)^{-q+1} R_q(\lambda) = \left(\frac{1-e^{-x}}{x} \right)^{-q+1} \sum_{j=1}^{\infty} (xj)^{-q} e^{-xj} x \rightarrow \int_0^{\infty} z^{-q} e^{-z} dz = (-q)!. \quad (\text{D.5})$$

For $q = 0$ and $q = 1$

$$R_0(\lambda) = \lambda/(1-\lambda), \quad R_1(\lambda) = -\ln(1-\lambda). \quad (\text{D.6})$$

We can now deduce the following results with respect to the inverse function

$$\lambda = \lambda_q(R) \quad (\text{D.7})$$

defined by (D.1):

For $0 \leq q \leq 1$, λ is a continuous function of R for $R \geq 0$ and increases monotonically from $\lambda = 0$ to $\lambda = 1$ in the interval $(0, \infty)$, where the derivative exists and decreases monotonically from 1 to 0

For $1 < q \leq 2$, the same is true for the interval $0 \leq R \leq R_q(1)$. Regarding the second derivative

$$\frac{d^2\lambda}{dR^2} = -\frac{\lambda(R_{q-2} - R_{q-1})}{R_{q-1}^3}, \quad (\text{D.8})$$

the following cases are to be distinguished:

$$1 < q < \frac{3}{2}: \quad \frac{d^2\lambda}{dR^2} = 0 \quad (R = R_q(1)), \quad (\text{D.9})$$

$$q = \frac{3}{2}: \quad \frac{d^2\lambda}{dR^2} = -\frac{1}{2\pi} \quad (R = R_q(1)), \quad (\text{D.10})$$

$$\frac{3}{2} < q \leq 2: \quad \frac{d^2\lambda}{dR^2} \rightarrow -\infty \quad (R \rightarrow R_q(1)). \quad (\text{D.11})$$

For $q > 2$, $d\lambda/dR$ is always positive for $0 \leq R \leq R_q(1)$.

E. THE LIMITING CASE OF CONSTANT ν

THEOREM 1. *Let N and each a_v tend to infinity with constant ν ,*

$$\nu \equiv N \left/ \prod_{v=1}^w a_v^{2/\alpha} \right., \quad (\text{E.1})$$

$$\text{then, with } q = w/\alpha, \quad \nu \{(\alpha^{-1})!\}^{-w} (M/T)^q = R_q(\lambda). \quad (\text{E.2})$$

For $q \leq 1$ (case A), (E.2) defines λ for every T . For $q > 1$ (case B), (E.2) is only valid for $T \geq T_0$, where T_0 is given by

$$\nu \{(\alpha^{-1})!\}^{-w} (M/T_0)^q = R_q(1), \quad (\text{E.3})$$

$$\text{whereas for } T \leq T_0 \quad \lambda \equiv 1. \quad (\text{E.4})$$

The proof is as follows. Denoting the left-hand side of (E.2) by $u(T)$, we transform (A.6) into

$$u(T) = \sum_j \lambda^j j^{-q} \prod_{v=1}^w [h_\alpha(x_v j) + \{(\alpha^{-1})!\}^{-1} (x_v j)^{1/\alpha}], \quad (\text{E.5})$$

i.e. carrying out the multiplication,

$$u(T) = A_0 + A_w + \sum_{v=1}^w A_v, \quad (\text{E.6})$$

in which each term A_v contains the factor $\{(\alpha^{-1})!\}^{-v}$,

$$A_0 = \sum_j \lambda^j j^{-q} \prod_{v=1}^w h_\alpha(x_v j) \quad (\text{E.7})$$

$$\text{and} \quad A_w = \{(\alpha^{-1})!\}^{-w} \left(\frac{M}{T}\right)^q \frac{\lambda}{1-\lambda} \prod_{v=1}^w a_v^{-2/\alpha}. \quad (\text{E.8})$$

Multiplying (E.6) by $(1-\lambda)$ we obtain

$$\{u(T) - A_0\}(1-\lambda) = A_w(1-\lambda) + \sum_{v=1}^{w-1} A_v(1-\lambda), \quad (\text{E.9})$$

where $A_v(1-\lambda)$ is not negative and, because of $h_\alpha < 1$, is minorant to

$$\{(\alpha^{-1})!\}^{-v} (M/T)^{v/\alpha} \{\sum (a_{n_1} \dots a_{n_v})^{-2/\alpha}\} (1-\lambda) R_{(w-v)/\alpha}(\lambda), \quad (\text{E.10})$$

where Σ consists of $\binom{w}{v}$ terms. $A_w(1-\lambda)$ and the majorant (E. 10) are finite, even for $\lambda = 1$ (see also (D. 3)). Then in the limiting case of infinite N and a_v , $A_w(1-\lambda)$ and the expression (E. 10) tend to zero. At the same time A_0 tends to $R_q(\lambda)$ because of (C. 4). So, in the limiting case, (E. 9) transforms into

$$[\nu\{(\alpha^{-1})!\}^{-w}(M/T)^q - R_q(\lambda)](1-\lambda) = 0. \quad (\text{E. 11})$$

We now have to distinguish two cases:

Case A, $q \leq 1$. As $R_q(\lambda)$ increases monotonically from $0(\lambda = 0)$ to $+\infty(\lambda = 1)$, (E. 11) is satisfied at each temperature T by exactly one value of λ between 0 and 1. So (E. 2) defines λ as a continuous function of T , decreasing monotonically from $\lambda = 1(T = 0)$ to $\lambda = 0(T = \infty)$.

Case B, $q > 1$. Since $R_q(\lambda)$ does not exceed the finite value $R_q(1)$, the first factor of (E. 11) is positive for $T < T_0$, T_0 given by (E. 3), and so $\lambda = 1$. For $T \geq T_0$, λ is again defined by (E. 2). This proves the theorem.

Clearly, in case B the two functions join continuously at $T = T_0$. As to the character of the transition, we have, in view of (D. 7), for $T \geq T_0$,

$$\lambda = \lambda_q\{u(T)\}, \quad (\text{E. 12})$$

$$\frac{d\lambda}{dT} = \frac{d\lambda}{dR} \frac{du}{dT}, \quad (\text{E. 13})$$

$$\frac{d^2\lambda}{dT^2} = \frac{d^2\lambda}{dR^2} \left(\frac{du}{dT}\right)^2 + \frac{d\lambda}{dR} \frac{d^2u}{dT^2}. \quad (\text{E. 14})$$

With the aid of §D, this leads to the following résumé: *In case A ($q \leq 1$), $\lambda(T)$ has no transition point, while in case B ($q > 1$) a transition point T_0 is given by (E. 3) and is characterized by a discontinuity in the third or a higher derivative for $1 < q < \frac{3}{2}$, by an infinite (finite) discontinuity in the second derivative for $\frac{3}{2} < q \leq 2$ ($q = \frac{3}{2}$), and by a finite discontinuity in the first derivative for $q > 2$.*

THEOREM 2. *In the limiting case of theorem 1 for $T < T_0$ (case B), the number N_0 of particles in the lowest state is also infinite, in such a way that N_0/N is finite and*

$$1 - \left(\frac{N_0}{N}\right) = \left(\frac{T}{T_0}\right)^q. \quad (\text{E. 15})$$

The proof is as follows. According to (1.3), with finite N and a_v , we can write (E. 8) in the form

$$A_w = \left(\frac{N_0}{N}\right) u(T). \quad (\text{E. 16})$$

We prove that, in the limiting case for $T < T_0$ ($\lambda \rightarrow 1$),

$$A_w \rightarrow u(T) - R_q(1), \quad (\text{E. 17})$$

which, in view of (E. 3), immediately leads to (E. 15). We have

$$u(T) - A_0 > u(T) - R_q(\lambda) > 0, \quad (\text{E. 18})$$

the last inequality being that stated after (E. 11). With $N \rightarrow \infty$, $u(T) - A_0$ tends to $u(T) - R_q(1)$, and so does $A_1 + \dots + A_w$, because of (E. 6). We shall show that this

happens in such a way that $A_1 + \dots + A_{w-1}$ tends to zero. This will prove (E.17). With all a_y greater than some large r , let λ tend to 1 in such a way that A_w is kept finite, i.e. because of (E.8),

$$1 - \lambda \sim r^{-2q}. \quad (\text{E.19})$$

Then, because of (E.10), for $1 \leq v \leq w$, A_v is minorant to

$$\text{constant } r^{-2v/\alpha} R_{(w-v)/\alpha}(\lambda). \quad (\text{E.20})$$

With $v < w - \alpha$, the series in λ converges even for $\lambda = 1$, and so (E.20) tends to zero with increasing r . With $v = w - \alpha$, (E.20) is of the order of $r^{-2v/\alpha} \ln r$ (see (D.6)), and therefore vanishes as $r \rightarrow \infty$. Finally, with $w - \alpha < v < w$, (E.20) is of the order of $r^{2(w-v)/\alpha(1-w/\alpha)}$ (see (D.3)), and hence tends to zero in the limiting case, since we have $w/\alpha > 1$ in case B. Thus, with finite A_w , $A_1 + \dots + A_{w-1}$ vanishes, which proves the theorem. [If we had taken care to keep one of the terms A_v finite, then A_w would have tended to infinity and, all the A_v being positive, we should have come into conflict with the fact that $u(T) - A_0$ is finite.]

THEOREM 3. *In the limiting case of theorem 1, in case A as well as in case B, for all temperatures $T > 0$,*

$$\bar{\epsilon} = \frac{kT}{\nu} \{(\alpha^{-1})!\}^w q(T/M)^q R_{1+q}(\lambda) \quad (\text{E.21})$$

tends to $w\{(\alpha^{-1})!\}^{w-1} \lambda^q j^{-1-q}$, the theorem is proved.

For all temperatures in case A, and for $T \geq T_0$ in case B, theorems 1 and 2 combine to give

$$\bar{\epsilon} = qkTR_{q+1}(\lambda)/R_q(\lambda), \quad (\text{E.25})$$

so that the value of $\bar{\epsilon}$ is smaller than the 'classical' expression

$$\bar{\epsilon}^* = qkT, \quad (\text{E.26})$$

the quotient in (E.25) being smaller than 1. Only for $\lambda \rightarrow 0$ ($T \rightarrow \infty$) does this quotient tend to unity. It can be shown from the behaviour of $\lambda(T)$, that, in both cases A and B, as $T \rightarrow 0$, $\bar{\epsilon}(T)$ tends to zero like T^{1+q} and $d\bar{\epsilon}/dT$ like T^q . For $0 < T \leq T_0$ in case B,

(E.21) is valid with $\lambda \equiv 1$. At the transition point, $\bar{\epsilon}(T)$ is continuous since this is true for $\lambda(T)$. Calculating the specific heat $d\bar{\epsilon}/dT$ from

$$\frac{d\bar{\epsilon}}{dT} = \frac{\partial \bar{\epsilon}}{\partial T} + \frac{\partial \bar{\epsilon}}{\partial \lambda} \frac{d\lambda}{dT}, \quad (\text{E.27})$$

and $d^2\bar{\epsilon}/dT^2$ in a similar manner, one finds, on the basis of the résumé preceding theorem 2, the following results:

In case B, $\bar{\epsilon}(T)$ is continuous at the transition point $T = T_0$, with a continuous second derivative for $1 < q < \frac{3}{2}$, with an infinite (finite) discontinuity in the second derivative for $\frac{3}{2} < q \leq 2$ ($q = \frac{3}{2}$), and with a finite discontinuity in the specific heat $d\bar{\epsilon}/dT$ for $q > 2$. In this last case there is a λ -point.

From (E.25) it can be seen that at $T = T_0$ ($\lambda = 1$), $\bar{\epsilon}$ is a constant fraction of $\bar{\epsilon}^*$, so that T_0 can be used as a characteristic temperature in the sense in which this word is used, for example, in the Debye theory of the specific heat of solids.

It should be noted that the results for $\alpha = 2$, $w = 3$ (three-dimensional box) in the limiting case considered in this section are not affected by replacing the expression (A.3), i.e. (2.1) for ϵ_i by (2.18). This introduces only slight modifications into the functions $G_2(x)$, etc.

We quote the results in the following special cases: $R_q(1) = 2.612, 1.645, 1.341, 1.202, 1.0823$ for $q = \frac{3}{2}, 2, \frac{5}{2}, 3$ and 4 respectively (Jahnke & Emde 1945).

For the box, $\alpha = 2$, $M = \hbar^2/(8mk)$, $(\alpha^{-1})! = \frac{1}{2}\sqrt{\pi}$. A transition temperature occurs with $w = 3$ (see (2.5)) but not with $w = 2$ or $w = 1$ (see (3.1) or (3.3)).

For the oscillator field, $\alpha = 1$, $M = 3\hbar^2/(4\pi^2mk)$, $(\alpha^{-1})! = 1$. With constant ν , transition points occur for $w = 3$ (see (4.12)) and $w = 2$ (see §5) but not for $w = 1$ (see (5.2)).

F. TWO OTHER LIMITING CASES

Now let N and each a_v tend to infinity with constant 'density' n , $n = N/\prod_v a_v$. Then, with $1 \leq \alpha < 2$, $\nu \rightarrow 0$ (with $\alpha = 2$, $\nu \equiv n$), which leads to $T_0 \rightarrow 0$. So with fixed positive T , (E.2) and (E.25) hold for ν sufficiently small. Because of (E.2), $R_q(\lambda) \rightarrow 0$, and so $\lambda \rightarrow 0$ and $\bar{\epsilon}$, given by (E.25), tends to $\bar{\epsilon}^*$. This result can also be stated in terms of a quantity b defined by

$$b = T \prod_{v=1}^w a_v^{\alpha/w(2/\alpha-1)}. \quad (\text{F.1})$$

Reasonings, similar to those of the preceding sections, then lead to the statement that, in the limiting case with constant n , in case A and, for $b \geq b_0$, in case B,

$$n\{(\alpha^{-1})!\}^{-w}(M/b)^q = R_q(\lambda), \quad (\text{F.2})$$

where b_0 is given by (F.2) with $\lambda = 1$, whereas $\lambda \equiv 1$ for $b \leq b_0$ (analogous to theorem 1). Analogous to theorem 3, we now have for all positive b

$$\frac{\bar{\epsilon}}{kT} = \frac{1}{n} \{(\alpha^{-1})!\}^w q (b/M)^q R_{q+1}(\lambda), \quad (\text{F.3})$$

and (E.25) holds in the region of validity of (F.2). With fixed positive T , we are always in this region, since then $b \rightarrow \infty$.

Finally, we consider the limiting case $N \rightarrow \infty$ where all the a_v are assumed to be bounded, $r < a_v < R$, with finite r and R . It then follows from (A.6) that $\lambda \equiv 1$ for all T . Since now $x_{vj} > Mj/R^2T \geq M/R^2T$, $G_\alpha(x_{vj})$ does not exceed $G_\alpha(M/R^2T)$ and

$$-G'_\alpha(x_{vj}) < \text{constant } e^{-M/R^2T}, \quad (\text{F.4})$$

as can be seen from (A.14). It can then be proved from (A.7) that the total energy $N\bar{\epsilon}$ is bounded by a series which converges even for $\lambda = 1$, and is therefore finite in the limiting case $N \rightarrow \infty$. The number N_0 of particles in the lowest state is equal to $\sum_j \lambda^j$ (see (1.3)) and tends to infinity with $N \rightarrow \infty$; but with finite a_v , $N - N_0$ remains finite. The given series for N_0 combines with (A.6) to give a series for $N - N_0$ (with finite N). Since $G_\alpha(x_{vj})$ is always minorant to $G_1(x_{vj})$ which is equal to $(1 - e^{-x_{vj}})^{-1}$, and since now $x_v > M/R^2T$, this series can be proved to converge, even for $\lambda = 1$, which implies that $N - N_0$ is finite.

G. CALCULATION OF $\bar{\epsilon}$ WITH THE CONTINUOUS SPECTRUM APPROXIMATION

Let us assume the energy levels to be given by

$$e_{s_1, \dots, s_w} = \frac{Mk}{a^2} \sum_{v=1}^w$$

Representing the energy states by s_v , we proceed to the continuous spectrum with $\sum_v s_v^\alpha \leq \rho^\alpha$ is $\rho^w I$, where I is

$$I =$$

With $s_v^\alpha = t_v$, I transforms into

$$I = \int_{t_1+\dots+t_w \leq 1} \dots \int \alpha^{-w} dt_1 \dots dt_w$$

So, for the number of states in the energy interval $(\epsilon, \epsilon + d\epsilon)$, we obtain

$$d(\rho^w I) = wI\rho^{w-1}d\rho = \frac{1}{(q-1)!} \{(\alpha^{-1})!\}^w \left(\frac{a^2}{Mk}\right)^q \epsilon^{q-1} d\epsilon.$$

Of these states, the fraction

$$\frac{dN}{d(\rho^w I)} = (\lambda^{-1} e^{\epsilon/kT} - 1)^{-1}$$

is occupied (see alternative form (1.1)). The normalization yields

$$N = \int dN = \{(\alpha^{-1})!\}^w (\alpha^2 T/M)^q R_q(\lambda). \quad (\text{G.6})$$

The total energy is

$$E = \int \epsilon dN = qkT \{(\alpha^{-1})!\}^w (\alpha^2 T/M)^q R_{q+1}(\lambda) \quad (\text{G.7})$$

and

$$\bar{\epsilon} = E/N = qkT R_{q+1}(\lambda)/R_q(\lambda). \quad (\text{G.8})$$

In the case $q > 1$, however, $R_q(\lambda)$ does not exceed the finite value $R_q(1)$ (in view of (G. 5), λ should not exceed 1), and so (G. 6) is only valid for $T \geq T_0$, where T_0 can be obtained from (G. 6) with $\lambda = 1$. Obviously, (G. 6) and (G. 8) do not define λ and $\bar{\epsilon}$ for $T < T_0$ (these equations can be compared with (E. 2) and (E. 25) with the difference that, in the above reasoning, N and a are finite). This indefiniteness has already been pointed out by London (1938*b*) in the case of the box ($q = \frac{3}{2}$) and is caused by the substitution of a sum as in (A. 1) by an integral as (G. 6). The discrepancy between (A. 1), which defines λ for each T , and (G. 6) can be removed by a slight modification of the above approximation, i.e. by assuming the lowest level(s) to be discrete, as London did in the box problem.

The rigorous method of the preceding sections, which is a modification and generalization of that given by Fowler & Jones (1938), has been applied here to all the cases which until now have been treated by the continuous spectrum approximation (London 1938*b*, 1939; cf. also Mayer & Mayer 1946). It may, however, be that new problems will arise, more general than those described by (A. 3), which cannot be solved by the rigorous method, but could be tackled by the continuous spectrum approximation. Then London's modification of the latter method should be chosen, since it avoids discrepancies such as those described above which are sometimes encountered in the approximation method.

We wish to express our most sincere thanks to Professor E. A. Guggenheim of Reading University for suggesting this problem to us, for numerous helpful suggestions concerning this paper and for constant encouragement.

REFERENCES

- Einstein, A. 1925 *S.B. preuss. Akad. Wiss. Ber.* 3, pp. 18–25.
 Fowler, R. H. & Jones, H. 1938 *Proc. Camb. Phil. Soc.* 34, 573.
 de Groot, S. R. & ten Seldam, C. A. 1949 *Physica*, 15, 671.
 Jahnke, E. & Emde, F. 1945 *Tables of functions*, p. 273. New York: Dover Publications, 4th ed.
 Kaempffer, F. 1949 *Z. Phys.* 125, 359.
 Leibfried, G. 1947 *Z. Naturforsch.* 2A, 305.
 London, F. 1938*a*. *Nature*, 141, 643.
 London, F. 1938*b* *Phys. Rev.* 54, 947.
 London, F. 1939 *J. Phys. Chem.* 43, 49.
 Mayer, J. E. & Mayer, M. G. 1946 *Statistical mechanics*. New York: John Wiley & Sons, Inc.
 Schubert, G. 1946 *Z. Naturforsch.* 1, 113.
 Sommerfeld, A. 1946 *Z. Naturforsch.* 1, 120.
 Uhlenbeck, G. E. 1927 *Over statistische methoden in de theorie der quanta*, p. 70. The Hague Martinus Nijhoff.
 Wergeland, H. 1944 *K. norske vidensk. Selsk. Forh.* 17 (13), 51, 17 (15), 63.

The Bell Telephone Laboratories—an example of an institute of creative technology

BY MERVIN J. KELLY

Bell Telephone Laboratories, New York, N.Y.

(Lecture delivered 23 March 1950—Received 23 May 1950)

[Plates 4 to 6]

INTRODUCTION

The part that science and technology have played in influencing the economic, social and political patterns of western society and in enriching the lives of its people has steadily increased during the last century. The scope and character of this influence have varied widely from country to country. Traditions, mores, maturity, size, patterns of education, and many other factors have been elements in bringing about the variations.

The influence has probably been most profound in the United States, principally, I believe, because of its youth, size, and patterns of education. Beginning some four or five decades ago, my country has been transformed at an increasingly rapid tempo from primarily an agricultural society to predominantly an industrial one under the driving force of an expanding body of science and technology. So completely have they dominated the pattern of our growth that when the man in the street speaks of 'progress', he usually means scientific and technological progress.

Until the beginning of this century, the work in applying new scientific knowledge to new facilities and instrumentalities for society was quite a hit and miss process. The inventor, having little or no direct contact with pure science, took the first steps whose end product was the inventor's model. Then the engineer, who at that time was largely a graduate from the drafting board or from the shop, reduced the inventor's model to the design of a new product for manufacture. While progress was made, the procedure was slow, inefficient, and the intervals of time were quite long between the availability of new scientific knowledge and the appearance of new products made possible by it.

The break from this pattern began at the turn of the century with the appearance in industry of men trained in the scientific method of research. They were the pioneers of industrial research and established industry's first laboratories of applied science. Dr Jewett of the Bell System and Dr Whitney of the General Electric Company were among the first. Other men with similar training followed them, initially at a slow rate, but, as they proved their value, the number rapidly grew. We now have some 2500 industrial research laboratories. They vary in size from the little fellow with less than fifty employees to the giant with more than a thousand. As they have grown in size, aged, and learned through experience, their quality has steadily improved.

To keep pace with the evolution of its research laboratory and take advantage of the opportunities accruing from the adoption of the scientist and his methods, the engineering organization of industry has undergone major change. Its relatively simple operation, in the last century, of transforming the inventor's model into a design for manufacture, performed largely by empirical methods, has now expanded into many successive interlaced operations. Each, as it has matured, employs more of the scientific method and of fundamental analysis in the solution of its problems.

The applied science laboratory and these engineering functions have been integrated in various ways into the organization of our industrial corporations. Whatever the organization may be, the work is a continuously flowing scientific and technical operation that begins at the forefront of the areas of pure science that are significant to the industry, and proceeds through successive steps of research, development, design, and engineering for manufacture of the new product or facility.

There has been so much emphasis on industrial research and mass-production methods in my country, that even our well-informed public is not sufficiently aware of the necessary and most important chain of events that lies between the initial step of basic research and the terminal operation of manufacture. In order to stress the continuity of procedures from research to engineering of product into manufacture and to emphasize their real unity, I speak of them as the single entity 'organized creative technology'.

HISTORICAL

I am using the Bell Telephone Laboratories and its operations as an exemplification of this unity. But first I must make sure that all of you are familiar with us.

The Laboratories is an organization of about 5700 scientists, engineers, and supporting staff—2200 are scientists and engineers. It serves the Bell System—an operating or service organization which provides nation-wide telephone service for the United States. Some 34 million telephones are owned and operated by the Bell Companies and are interconnected with some 7 million additional telephones operated by a large number of relatively small independent companies.

The Laboratories provides the Bell System with new communications technology and the information for its implementation with new systems and facilities. Its work begins at the forefront of applied science, extends across the entire scope of creative technology through research, development, design, and engineering. The end products of its programmes are the designs and specifications for manufacture of the technical facilities that it has created and the engineering practices to be followed in their operation and maintenance.

Since the earliest days of the telephone, the Bell System has been responsible for its own technology and for the manufacture of the major portion of its technical facilities. The Western Electric Company, a subsidiary, has always been the manufacturer.

The Bell System is unique among organizations providing service in its responsibility for its own technology and for the manufacture of its technical facilities. The service of telephony differs fundamentally from that of others, such as power, gas

and water, in its technical complexity, its essential national integration, and in its two-way character. The telephone system of the United States can be viewed as a single, integrated, highly technical machine in which electrical currents that are very small and complex in wave form are sent from any one of more than 40 million points to any one of all the others. The intensity (amplitude) and wave form of these currents must be accurately controlled and, because of their extremely low value, interfering electrical currents of a great variety of kinds must be either excluded from circulating in the network or maintained at a level well beneath that of the 'intelligence' currents.

This integration of technology, manufacture, and operation has permitted the rapid application of new scientific knowledge to new and improved telephone services and the realization of sound engineering and economic solutions to the large variety of the complex problems of a nation-wide telephone service. This has been a major element in providing the American public with the most complete telephone service of the world and at a cost, in terms of the income of its people, so low that we make use of telephony in our daily economic and social life to a greater extent than any other people.

It was not until 1925 that all the functions of creative technology were consolidated into a single organization. At that time the Bell Telephone Laboratories was formed as a subsidiary corporation. While corporate and organizational unification occurred only 25 years ago, our operations in creative technology had been under development during the preceding 20 years. Dr Jewett, the first president of our Laboratories and one of the first scientists engaged in Bell System work in technology, exerted an increasing influence on all sectors of our creative technology independent of its organizational pattern.

We have, therefore, a background of almost a half-century of experience and a professional staff whose age distribution in the years from 25 to 65 is now substantially uniform. Through trial and error with continuous appraisal of different ways for organization of work, types of men and their training, and of housing and facilities, our present patterns of work, of men and their training, and our housing and facilities have evolved.

In the half-century in which our Laboratories has evolved, the physical sciences underlying telecommunications have made greater progress in advancing knowledge of the laws of nature and the structure of matter than in any equal period of history. The increase in fundamental knowledge has been at a rate so rapid that we have been faced with an increasingly complex problem in maintaining a close linkage between the forefront of our applied research and that of pure science.

It was during this same period that the graduate schools of our country entered into full participation in research in the physical sciences as an equal partner with the older society of Europe. At the beginning of the century there were only a few of our universities in which research at the forefront was in progress.

From this small beginning there has been an increasingly rapid expansion in the volume of research and in the number of universities and institutes of technology participating, until to-day there are active and effective programmes in research in the physical sciences in almost a hundred universities and institutes of technology.

This expansion has been accompanied by the training in each year of a greater number of men at the graduate level in the physical sciences and in the methods of scientific research.

Our Laboratories, therefore, grew in size and matured in the scope and character of its work during the period of rapid expansion in research in the physical sciences and in the number of men capable of working in the most modern methods of scientific research. This presented us with a most difficult problem of maintaining a leadership and a professional staff of adequate competence and training to establish programmes based on the most recent scientific knowledge and to carry them out with methods and instrumentation that kept pace with the progress of pure research.

THE ORGANIZATION OF OUR WORK

To present the Laboratories as a mature institute of creative technology, I shall consider its operations under three general headings.

The first includes all of the research and fundamental development. This is our non-scheduled area of work. It provides the reservoir of completely new knowledge, principles, materials, methods, and art that are essential for the development of new communications systems and facilities. Arising in the course of the work and forming an important part of it is a large fraction of those creative concepts of our Laboratories that constitutes invention.

Its programmes extend across all sectors of science that may contribute to the advancement of the communication art. They are carried out in sufficient volume that there is the minimum reasonable time lag between an advance in pure science and our realization of its contribution to our reservoir of new knowledge.

The second we call 'systems engineering'. Its major responsibility is the determination of the new specific systems and facilities development projects—their operational and economic objectives and the broad technical plan to be followed. 'Systems engineering' controls and guides the use of the new knowledge obtained from the research and fundamental development programmes in the creation of new telephone services and the improvement and lowering of cost of services already established. In determining the new development projects, 'systems engineering' considers the content of the reservoir of new knowledge awaiting application and the opportunities for its use in the interest of the telephone user. The projects that are activated are those that give greatest promise for user benefit. It attempts to insure that the technical objectives of the development projects undertaken can be realized within the framework of the new knowledge available in the reservoir and present engineering practice.

The third encompasses all specific development and design of new systems and facilities. The work is most carefully programmed in conformity with the plan established by the systems engineering studies. Our research and fundamental development programmes supply the new knowledge required in meeting the objectives of the new specific developments.

Before a new development project is activated, 'systems engineering' establishes well-considered performance and cost objectives. It determines the general technical

line of attack sufficiently well to insure that there is adequate fundamental knowledge available in our reservoir for its realization. The development project is carried out under the charter established by systems engineering. There are three well-ordered and successive steps in its performance. The end products are the specification of designs and materials for the manufacture of the system or facility, and the engineering practices to be followed in its operation and maintenance in service.

I shall now turn to a more detailed examination of the programmes of each of the three areas, their interrelationships, and the types and training of the men that we have found best for the work.

RESEARCH AND FUNDAMENTAL DEVELOPMENT

Researches in the sectors of physics, chemistry, and mathematics that give promise of contribution to advances in telephony are carried out in the forefront of the first area. Solid state physics, including magnetism, piezo-electricity, dielectrics, and semiconductors; physical electronics, electron dynamics; acoustics; electromagnetics; mathematics; organic and physical chemistry of synthetic plastics and rubber; corrosion chemistry; physical metallurgy; and fundamental mechanics are a typical but not comprehensive list of subjects.

The research area provides the coupling between the ever-advancing forefront of pure science and the forward march of our communications technology. The closer the coupling, the more completely will we keep our advancing technology in step with the progress of science.

Years of experience have taught us that the type and quality of men selected for our research, the environment that we provide, and the distance in their work that we ask them to penetrate beyond the forefront of creative technology are the most important factors in determining the closeness of coupling—the effectiveness.

Inspired and productive research in industry requires men of the same high quality as is required for distinguished pure research in our universities. We select our young men of research from among the most able and promising of the doctorate and post-doctorate of philosophy students of our graduate schools. We have no difficulty in attracting men of this type and quality.

They must be given freedoms that are equivalent to those of the research man in the university. This is, indeed, difficult in industry, but we are approaching that ideal. We give much attention to the maintenance of an atmosphere of freedom and an environment stimulating to scholarship and scientific research interest.

Our research staff make a practice of publishing the new scientific knowledge resulting from their programmes in the same manner as do the research men of the universities. In 1949 the members of our research staff published more than 200 scientific papers resulting from their research programmes. There are now seven treatises in preparation that will appear in book form in the Bell Telephone Laboratories-Van Nostrand series. Our scientists attend and participate in the meetings of their professional societies. They have a most active science seminar with programmes extending over a wide area of scientific interest. Fourteen

British scientists have been among the visiting speakers at the seminar during the past four years.

It is most important for the scientists to confine their efforts to the area of research. If they extend the area of their effort even to that of fundamental development (the area of work that immediately follows research) they tend to lose contact with the forefront of their field of scientific interest. In time, a considerable fraction will lose their productivity in research.

To make possible and to encourage this concentration of the attention of the men of science to research, we provide, through organization and stimulated association, an intimate tie between research and fundamental development, the next step in the chain of events from research to manufacture and use. In this way the programmes of research are taken over at a well-considered point by fundamental development, where they are extended and enlarged upon to supply the body of basic technology essential for the specific development and design of systems and facilities. While the fundamental development work is done in the best research tradition, it has a large content of the technologic, and economic considerations begin to be a factor in its programmes.

Staff Members for fundamental development are drawn from our research groups by selecting those having technologic and engineering aptitudes and interests who prefer to move into development and by recruitment from among the most promising of the graduate students of our schools of applied science, such as Massachusetts and California Institutes of Technology.

A typical example of a programme of this area will be helpful in making clear the distinctions between research and fundamental development and their inter-relationships. A few years ago members of our solid state physics research group focused their attention on the mechanisms of conduction in semiconductors, such as germanium and silicon. Important new knowledge of the conduction process acquired in the researches served as an inspiration and background for experiments with means and techniques for controlling the flow of the current in a semiconductor.

Out of this work came solid state amplifiers with characteristics quite similar to those of the thermionic vacuum tube amplifier. This is a discovery of major importance to communication technology. In mid-1948, after the scientists had rounded out their understanding of the physics of the amplification process, the Laboratories made public announcement of the solid state amplifier, which we called the 'transistor'. Concurrently, the scientists working on the project published the results of their researches, and have since published new findings and frankly discussed this new phenomenon with their peers.

In accord with our policy of concentrating the efforts of our scientists on research, we immediately formed a closely associated fundamental development group to acquire that body of technological knowledge essential to the development and design of transistors for the many specific communications applications that would certainly follow. They have interested themselves in such problems as the factors controlling the bandwidth of amplification; the noise figure; the amount of amplification possible per stage; energy levels of output; basic materials, processing and structure studies essential for controlled development and design of transistors for specific functions;

etc. Their work has been followed with sympathy and interest by the scientists in solid-state research. The fundamental development group has continuously consulted with the men of research on many of their problems. The contact of the scientists with the fundamental development work, as is to be expected, has not only been of great value to development but has stimulated further research work.

This pattern repeats itself again and again. A research programme is initiated. Then as new knowledge that gives promise of worth-while application is obtained, a fundamental development team is activated. It builds a background of basic technology under the watchful eye and with the consultative aid of the men of research.

Approximately 30 % of our professional staff work in the area of research and fundamental development. This provides adequate effort to maintain close coupling between our technology and science, and supplies a sufficient volume of new knowledge to feed into the programmes of our systems and facilities development area. That is to say, this division of effort gives good balance in our over-all programme.

SYSTEMS ENGINEERING

One of the principal responsibilities of systems engineering is technical planning and control. In the planning an appraisal is made of the various technical paths that can be followed in employing the new knowledge obtained by research and fundamental development in the specific development and design of new systems and facilities. The determination of the most effective use of new knowledge in the interest of the telephone user is the guiding principle of the planning studies. The most effective use may be the creation of new services, the improvement of the quality of existing services, the lowering of their cost, or some combination of these three. As the technology of communication has broadened and become more complex, the choice of the technical paths to be pursued in the instrumentation of the new technology has become increasingly difficult. It is this situation that has led to the evolution of the systems engineering function as a mechanism of guidance and control.

Systems engineering has intimate knowledge of the telephone plant and its operation and maintains close contact with the engineers of the operating organizations. The teamwork of operating engineers and our systems engineers makes available to the Laboratories in a most effective way the knowledge of the telephone system's needs and the opportunities for economy and improvement.

Systems engineering also maintains close association with our research and fundamental development work. It knows intimately the content of our new knowledge reservoir. It has equally effective liaison with the men and their work in the specific systems and facilities development area.

It integrates the knowledge from operations, research and fundamental development, and specific systems and facilities development. With this as a background, it makes exhaustive studies that appraise and programme development projects for new systems and facilities. Each study outlines the broad technical plan for a development, its objectives, and its economic and service worth. In many of

the studies it is recommended that no development be undertaken at the time. Administrative action is then taken with these studies as a guide to the selection and activation of the projects of our specific development programmes.

As the development organization proceeds with a project, systems engineering maintains close contact, continuously appraises the results, and amends the objectives and plans as required. Service trials are generally needed during the course of development. It organizes the trials in co-operation with operating engineers and participates in the tests and the evaluation of results. When the system is standardized and placed in manufacture, systems engineering follows service performance of first installations and coordinates the 'growing pains' engineering that is ever with us on new systems as they enter service. It finally participates in the evaluation of the service and economic worth that are experienced.

Typical examples of recent systems engineering studies that have led to development and standardization are: television transmission over coaxial cables, a broadband microwave radio repeated communication system, an automatic message accounting system, a mobile radio subscriber telephone system, and a new subscriber telephone set.

Systems engineering has another important responsibility. It recommends the levels of the various technical standards that are important elements in determining the quality and reliability of telephone service. Frequency bandwidth for voice transmission, noise levels on circuits, distortion and range of energy level of the speech currents, cross-talk levels, standards of protection against man-made and nature's interference with service are typical.

With advancing technology the levels of these various standards have gradually been raised. This has been reflected in improvements in quality, speed, and reliability of service. There must be a balance between service costs and rigour of these standards. It is systems engineering's responsibility to correlate with other organizations of the Bell System the factors involved in keeping standards and costs in balance so that a well-considered portion of economies available through advances in technology are used in raising quality and reliability and in increasing the speed of service.

Approximately 10% of our scientific and technical staff are allotted to systems engineering. Its staff members must supply a proper blending of competence and background in each of the three areas that it contacts: research and fundamental development, specific systems and facilities development, and operations. It is, therefore, largely made up of men drawn from these areas who have exhibited unusual talents in analysis and the objectivity so essential to their appraisal responsibility. For some of the more complex studies, the staff is augmented by borrowings of men having special knowledge or abilities from one or more of the three areas it coordinates.

SPECIFIC SYSTEMS AND FACILITIES DEVELOPMENT

The work of specific systems and facilities development is closely programmed. Its projects are organized in the patterns that the studies of systems engineering prescribe.



FIGURE 1 Aerial view of Murray Hill Unit of Bell Telephone Laboratories.

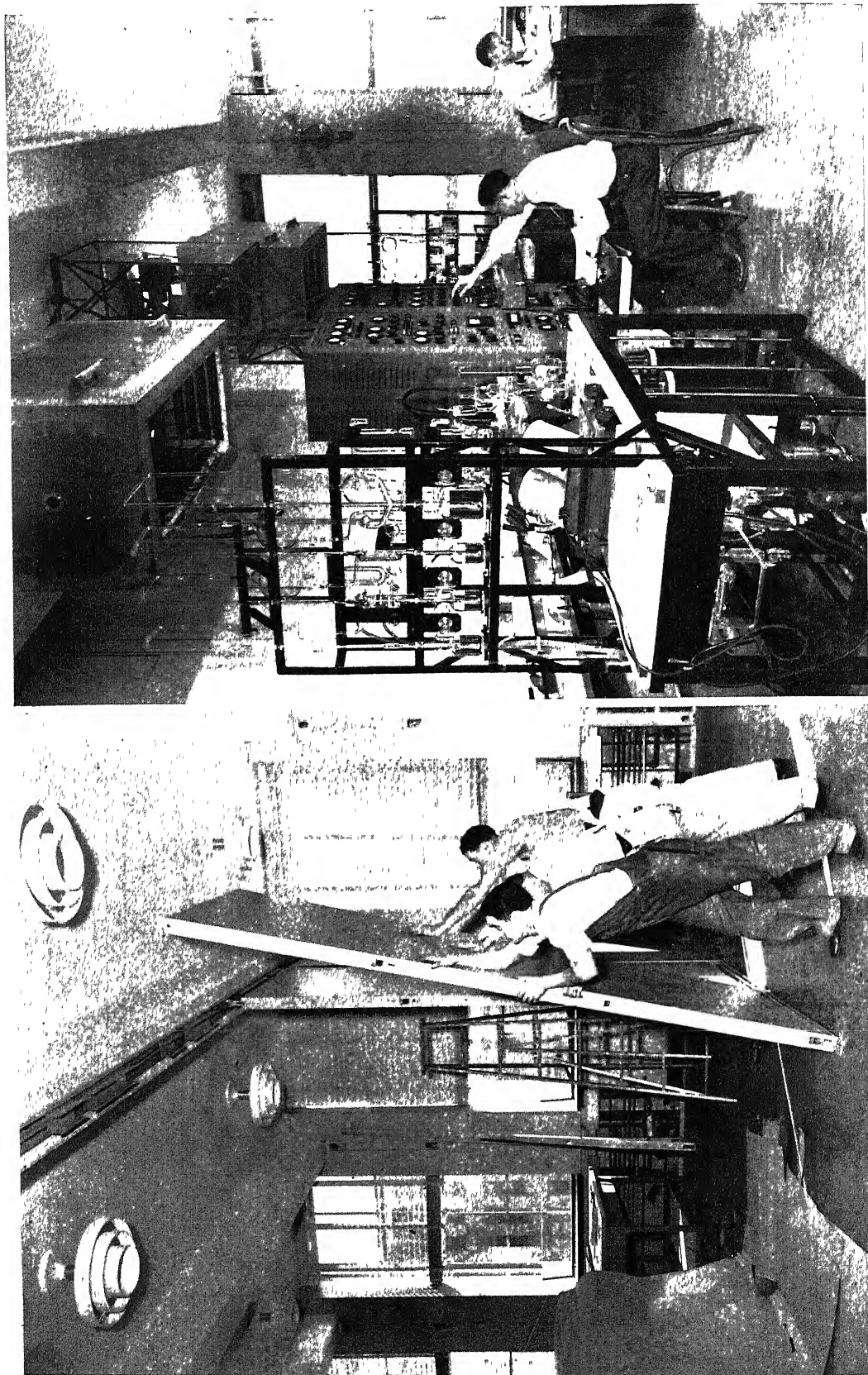


FIGURE 4. A typical physical research laboratory.

FIGURE 3. Placement of a panel of the removable metal partition.

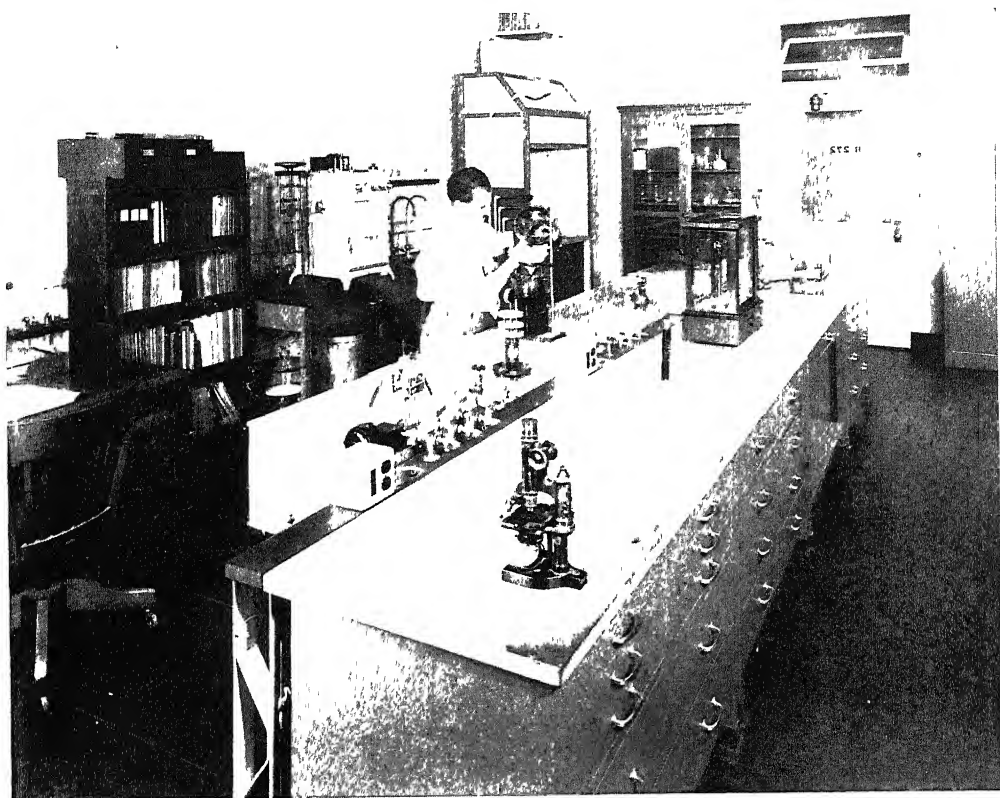


FIGURE 5. A typical chemical research laboratory.

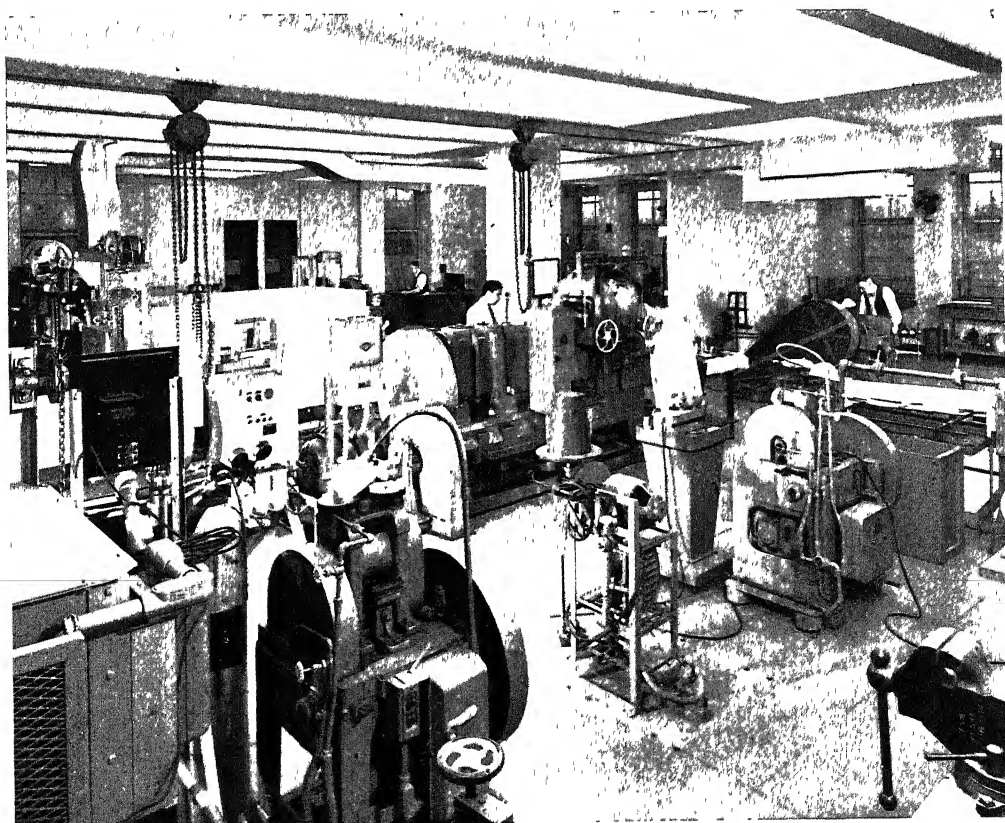


FIGURE 6. The metal fabrication unit of the metallurgical research laboratory.

The development, while a continuous operation, is done in three distinct stages. In the first, a laboratory model is evolved that meets the functional requirements of the systems engineering study. These are not always the same as initially established, for new information obtained during the course of development may well modify them. Systems engineering participates in their modification. Development is carried out within the framework of the new knowledge available in our new knowledge reservoir and of already established development and engineering practice. The model is thoroughly tested and proven for functional performance. Most intimate working contact is maintained between the specific development group and those of fundamental development throughout this stage.

In the next stage the design for manufacture and use is created. While retaining the functional performance of the laboratory model, it also meets the requirements of manufacture at lowest cost consistent with providing the specified service at lowest complete service cost. (The capital and depreciation charges that are directly related to manufacturing cost comprise only a portion of the complete cost of service.) In achieving these economic objectives in the design, the development group is the focal point of a closely integrated trifurcated team effort with manufacturing engineering of Western Electric as interpreters of, and important contributors to, design form for lowest manufacturing cost and with systems engineering as interpreters of operating requirements that are essential to lowest complete service cost.

The end product of stage two is a pre-production model. Here we are approaching final design. Western Electric produces a number varying from a few to a few hundred in their job shops and with preliminary tools. These models are placed in service under the close observation of the systems engineering and specific development groups of the Laboratories and the appropriate engineering groups of the operating companies.

The findings of the service tests and Western Electric's experience in producing the models are used as background for our final freezing of design and Western Electric's planning and tooling for quantity production. Design drawings and specifications are then prepared for manufacture, and engineering practices for the technical operations involved in supplying telephone service.

Some 60 % of the members of our professional staff are engaged in the development and design of systems and facilities. They are principally graduate engineers—electrical, mechanical, chemical, and metallurgical—from our technical schools and universities. The procedures of development and design have become increasingly fundamental and analytical. At the beginning of the century the work was largely empirical and done principally by graduates from the shop and drafting board.

We believe the maturity of science and fundamental technology that has been achieved offers opportunity for an even more fundamental and analytical approach in these final steps of creative technology. After much consideration of this problem during the post-war years, we concluded that a more fundamental approach can best be promoted if some of the engineers of specific development and design receive more fundamental training than they now obtain in the four- and five-year engineering curricula. We have, therefore, established within the Laboratories

a training course at the graduate school level in which selected engineering recruits receive added training in science and the technology of communications while on the job.

This completes the description of our work patterns and types of men involved in the successive steps of creative technology comprising research, fundamental development, systems engineering, and specific development and design. While programmes are systematically organized in this pattern and each member of the professional staff is assigned and works in that sector of creative technology for which his training, experience, and natural aptitudes best fit him, there exist important formally organized and informal team efforts within and across the boundary of each area.

So great is the complexity of applied science and technology of telecommunications that much specialization is required. The leaders of our programmes give a considerable portion of their time to the development of formal and informal co-operation. For example. While there are men with the necessary mathematical training and facility for the analytical work that normally is encountered in each of the areas of our work, the mathematicians of our research area devote at least one-fourth of their time to consultation and to aiding the analysts of the different areas of development in the solution of their problems. Such co-operation is informal and initiated by the men of development requiring the help.

Through the years our research and development leaders have developed patterns of informal co-operation and the habit of going to the expert, whether he be a mathematician, a metallurgist, an organic chemist, an electromagnetic propagation physicist, or an electron device specialist. This has made it possible to focus the full power of our organization on a particular problem. The organization's capacity for the solution of telecommunication problems is much greater than the sum of the capabilities of the individuals. Teamwork and co-operation are the important elements that bring about this amplification of strength.

The time that management gives to developing an environment favourable to teamwork and co-operation is most rewarding. It not only assures a high level of effectiveness in current programmes but, perhaps even more importantly, it fosters a continuing development of the members of staff and the institute as a living organism. It provides an organization of the flexibility and resilience so essential in meeting the challenges of a rapidly changing world of science and technology.

LABORATORY HOUSING, FACILITIES, AND SERVICES

Laboratory housing, facilities, and services is a topic that follows closely men, methods, and organization in its importance in the building of an efficient institute of creative technology. By the mid-1930's we fully recognized the limitation to our effectiveness imposed by the non-functional housing and old designs of facilities and services. A small committee was formed to determine the design of housing, facilities, and services best suited to effective laboratory operations.

In the later phases of the study the Bell System's architects were called in, and in co-operation with them we designed a functional structure to accommodate

1000 employees and with facilities and services complementary to the design. This laboratory was built on a 300-acre tract at Murray Hill, New Jersey, some 20 miles from New York City. It was occupied in 1941 and made a large contribution to the effectiveness of our war programmes. At the close of hostilities we made those modifications in design that experience indicated desirable and began the construction of an adjacent unit to house some 1600 employees. Its occupancy will be completed this year. The two units are connected by a passage-way at the second-floor level. An aerial view of the two buildings is shown in figure 1, plate 4.

The motif of the functional design of the buildings has been complete flexibility of room size with easily removable partitions for all interior walls. An institute of creative technology must be provided not only with space for laboratories and shops but also offices, conference rooms, drafting rooms, and clerical work. The structures provide two types of functional space, one for laboratories and shops and the other for offices, etc.

Laboratories and shop space comprise the main stems of the buildings. The stems are long, rectangular blocks of different lengths but of a uniform width of 54 ft. Our study indicated that we could most effectively meet the need for rooms of different sizes by having a repetitive 6 ft. module along the long axis of the rectangle. A corridor, 7 ft. wide, runs the length of the long axis of the rectangle. One side of the corridor is on the axis of the rectangle, thus providing rooms on either side of the corridor differing 7 ft. in depth. On one side of the corridor the depth is 27 ft. and 20 ft. on the other. This makes possible rooms 27 ft. in depth with widths that may be any integral multiple of 6 ft. on one side of the corridor and rooms 20 ft. in depth with widths that may be any integral multiple of 6 ft. on the other.

The repetitive 6 ft. unit was designed to include in its outer wall all of the facilities that might be needed in any laboratory space. Each 6 ft. unit was provided with a window for natural lighting and ventilation, artificial lighting fixtures, heating, ceiling inserts for support of apparatus, telephone, and access to fourteen types of services comprising various gases under pressure, electrical supply of different characteristics, as well as other services normally involved in telecommunications research and development.

Figure 2 is a diagram of two of the standard laboratory modules of 6 ft. and the distribution of service. This makes evident the manner in which the outside wall is used as the source of all services.

The adoption of the 6 ft. module with all services in the outer walls was predicated upon the use of a versatile system of wainscoting and partitioning to allow quick—often week-end—changes of work space. The entire building has a metal lining which is readily removable, and through which free access is given to the mechanical services.

The removable metal partition panels of the subdividing walls are the exact height to extend from the floor level to the ceiling beams, and are some 4 ft. wide. Each panel is 3 in. thick and consists of two layers of metal separated by mineral wool to make the panels fire-resistant and satisfactory in the elimination of sound transmission. In order to accommodate shelves, blackboards, or service piping supports, a keyholed device may be inserted between any two panels on 6 in.

vertical centres. The corridor walls are also made of the same subdivision panels, which are interchangeable with panels equipped with doors and transoms where these are needed.

Figure 3, plate 5, shows the placement of one of the removable metal partition panels in the construction of a subdividing wall. The metal lining of the outer wall is removed below the window level and the distribution piping for gases under pressure can be seen.

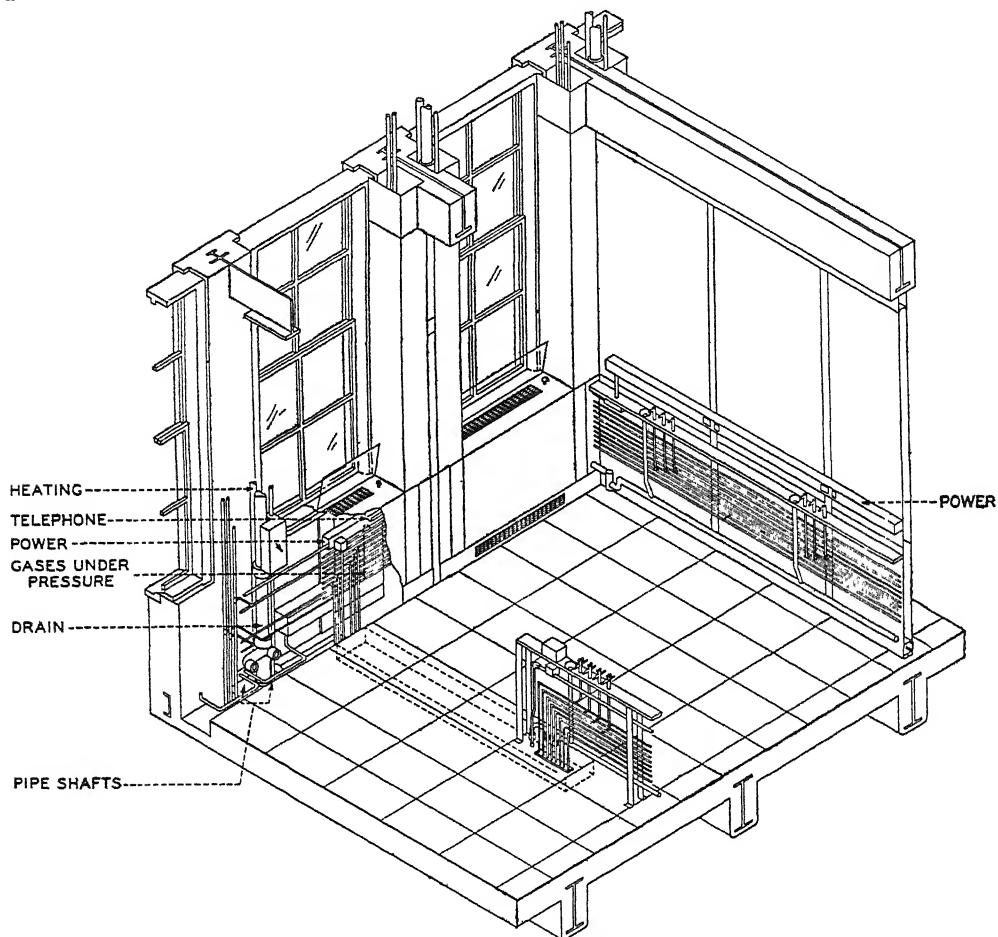


FIGURE 2. Diagram of two of the 6 ft. laboratory modules.

Laboratory furniture and facilities have been designed to complement the functional structure of the building. Every item of furniture and facilities, including even the chemical fume hood, has been made completely detachable and readily transportable. With the removable partitions and the furniture and facilities of special design, laboratory rooms of any required size can be readily constructed and equipped.

Figure 4, plate 5, shows the interior of a physical research laboratory in which gas-discharge studies are in progress. This laboratory is two modules or 12 ft. in width.

Figure 5, plate 6, shows the interior of a chemical research laboratory with an island bench. The services for the bench are extended to it from the outer wall by a trough under the floor that is integral with the building structure. This laboratory is three modules or 18 ft. in width.

Figure 6, plate 6, illustrates another feature of flexibility of the building design. This is the metal fabrication unit of our metallurgical laboratory. Because of the size of the machines, a very large space is required. This laboratory is eight modules or 48 ft. deep and 54 ft. wide. This width is obtained by placing the laboratory at one end of the rectangular block of a main stem and terminating the corridor eight modules from the end of the stem. This makes possible a laboratory extending across the full width of the structure. We have made similar use of the ends of the main stem rectangular areas in a number of situations to provide for laboratories of specially large area.

The flexibility of the laboratory space has made a large contribution to the operations of research and development. Rearrangement of laboratory rooms and facilities can be done expeditiously, with little loss of time, and at small expense. There is, therefore, no reluctance to rearrange space as the character of work changes. We can readily house in adjacent space the work on programmes where close relationships of the people involved are required. Development projects, by their nature, have small beginnings, and if successful expand as progress is made. The flexibility provided by our structure design makes possible the expansion of the working area of a development project as it matures.

The space for offices, conference rooms, drafting rooms, and clerical work is provided for in the wings that are at right angles to the main stems. This space does not require the large variety of special services that must be available to the laboratories. Economies were obtained by their segregation, which permits supplying those areas with only the limited services required by office space. The flexibility in room size has also been designed into these office and service wings. The same readily removable partitions are employed. The width of the module, however, has been increased from 6 to 9 ft., and the rooms on each side of the main corridor are 12 ft. in depth. The space can be subdivided into offices of different sizes, conference rooms, or large areas for drafting and clerical services.

Figure 7 is a diagram of a typical layout of laboratories in the main stem and of office, clerical, and drafting space in two service wings that are directly opposite each other across the width of the main stem.

DEVELOPMENT PROGRAMMES FOR THE MILITARY

The story of Bell Telephone Laboratories would not be complete without reference to our military development programmes. During the war we turned our attention almost completely to the creative technology of instrumentalities for warfare. New facilities for communications, radar, sonar, gun directors, etc. are typical of our areas of work.

At the close of hostilities, with military preparedness an important element of national policy, our Laboratories has continued to serve the nation in military creative technology.

The philosophy of work and organization of effort that have evolved in our technological services for an operating organization, the Bell System, makes us specially suited to serve the military, another operating organization. The pattern

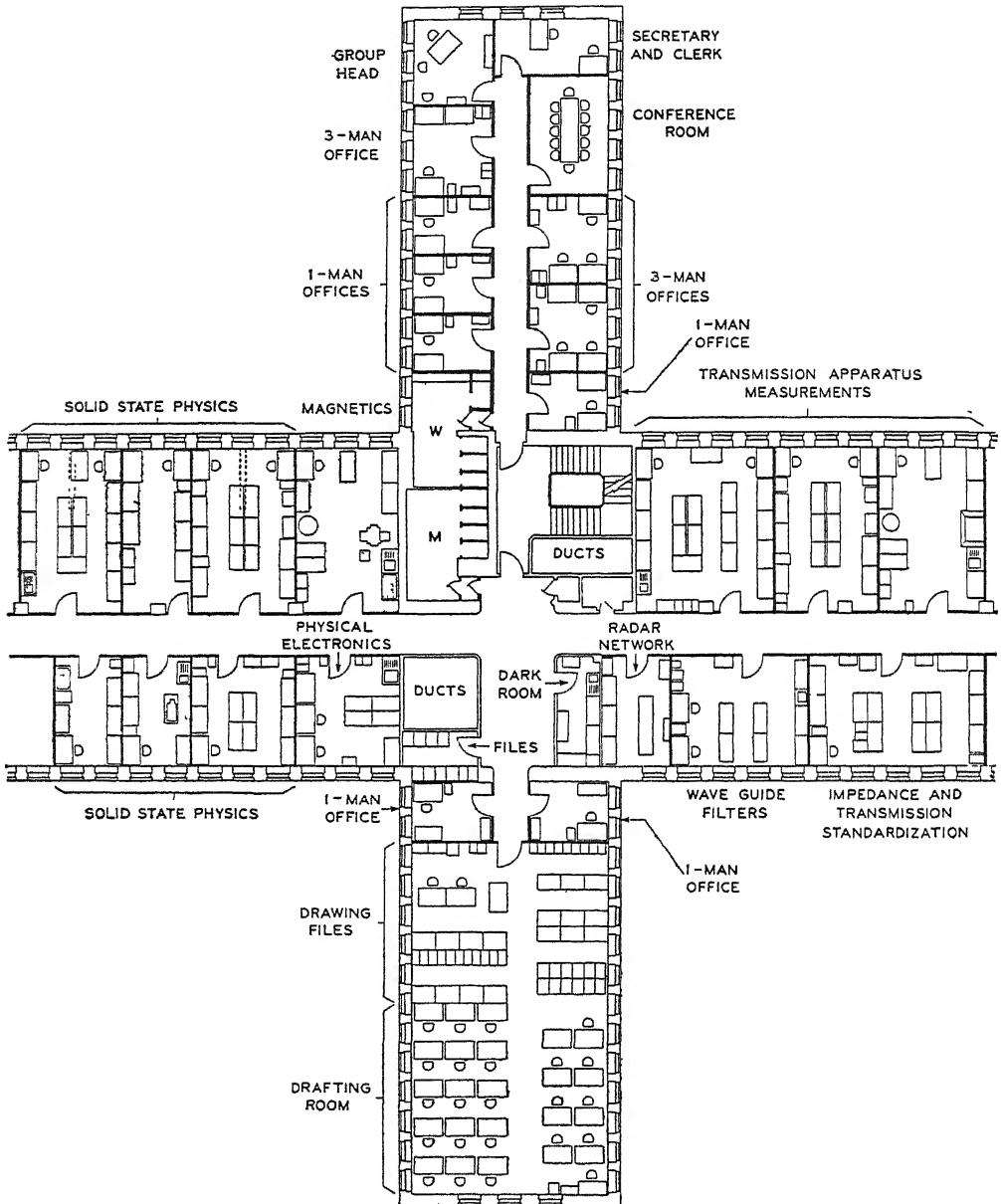


FIGURE 7. Diagram of a layout of a typical laboratory and office service area.

of our activities for the Bell System—a broad programme of research and fundamental development to obtain new knowledge of value to telephone service, the scientific determination of the best applications of the new knowledge in the

interest of service, and the development and design of new facilities best suited to service needs—is the pattern in which we are carrying out our work in creative technology for the military. The work on military systems proceeds in the same orderly manner through the successive steps from research to designs and specifications for manufacture and engineering practices for service which has proved so successful in the creation of new and improved systems for the services of telephony.

Studies in copolymerization

The evaluation of the kinetic coefficients for the system styrene-butyl acrylate

BY E. J. ARLMAN AND H. W. MELVILLE, F.R.S.

Chemistry Department, University of Birmingham.

Received 11 March 1950

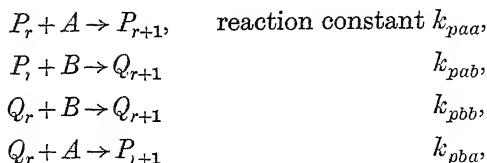
The mechanism of the copolymerization of styrene and butyl acrylate has been investigated in mixtures of these monomers using benzoyl peroxide as initiator. The peroxide was decomposed both thermally and photochemically. Experiments on the composition of the polymer as a function of the monomer composition have given the reactivity ratios styrene radical-styrene molecule to styrene radical-butyl acrylate molecule as 0.48 and butyl acrylate radical-butyl acrylate molecule to butyl acrylate radical to styrene molecule as 0.15. The kinetics of the copolymerization have been determined, and using the theory already elaborated for the treatment of such reactions, it has been shown by rate and molecular weight measurements that the styrene radicals react with the butyl acrylate radicals more quickly than the mean rate of reaction of styrene to styrene and butyl acrylate to butyl acrylate radicals. The coefficient ϕ defining this increased activity is not in this case constant and rises with increasing styrene content to a value of as much as 150. The initiation mechanism is also complicated in that there is no linear relation between the rate of initiation and monomer composition; instead, the rate drops rapidly when styrene is added to the butyl acrylate.

I. THE DETERMINATION OF MONOMER REACTIVITY RATIOS

INTRODUCTION

Until recently, in the study of addition copolymerization, attention has been mainly devoted to the relationship between composition of the copolymer and composition of the monomer mixture in which the copolymer is formed. The copolymer composition equation (Mayo & Lewis 1944; Alfrey & Goldfinger 1944) gives a quantitative description of this relationship and consequently makes it possible to state accurately the conditions for the preparation of copolymers of a given composition, of a desired chemical homogeneity or heterogeneity. According to this equation every monomer pair is characterized by two constants, the monomer reactivity ratios. These are the ratios of the reaction constants of both monomers with each of the radicals

present in the copolymerizing system. Calling the monomers A and B the radicals with an A -end P_r and with a B -end Q_r the propagation steps in copolymerization are



and the monomer reactivity ratios are defined by

$$r_1 = \frac{k_{paa}}{k_{pab}}, \quad r_2 = \frac{k_{pbb}}{k_{pba}}.$$

In order to elucidate as completely as possible the mechanism of copolymerization a knowledge of the monomer reactivity ratios is necessary but not sufficient, and it is essential to investigate also the initiation, termination and transfer steps. Particularly interesting features of such an investigation are the possibilities of determining the rate at which dissimilar radicals react with each other in the termination step and in checking whether a linear relationship exists between rate of initiation and composition. The theoretical basis of such an investigation may be found in the paper of Melville, Noble & Watson (1947), where the kinetics of the complete copolymerization system have been worked out on the general assumption that reactivity of polymer radicals is independent of size.

The systems styrene-methyl methacrylate and styrene-butyl acrylate have been investigated simultaneously along similar lines. It has already been reported (Melville & Valentine 1950) that in the system first mentioned dissimilar radicals react 14 times more quickly than would be expected from the mean reactivities of similar radicals. Thus if k_{taa} and k_{tbb} are the coefficients for the reaction of similar radicals the value of the crossed coefficient k_{tab} is given by $k_{tab} = 14(k_{taa}k_{tbb})^{\frac{1}{2}}$.

Moreover, rates of initiation using benzoyl peroxide as a photo-sensitizer vary in a peculiar way with monomer composition being lower than calculated for a linear dependence.

The same effects are found to exist even to a higher degree in the system styrene-butyl acrylate, which appeared also suitable for an investigation of this kind as the rate coefficients for both monomers are known and as the monomers readily copolymerize forming copolymers soluble in the reaction mixture. Unfortunately, the reactivity ratios of the monomers have not been determined, and this aspect of the problem forms the first section of this paper.

MATERIALS

The materials used were thoroughly purified.

Styrene (Abracene from Boake Roberts) was freed from inhibitors by washing repeatedly with 10 % sodium hydroxide solution and then with distilled water. After drying over anhydrous calcium chloride it was distilled *in vacuo* and stored in a dark bottle at a temperature of at least -20°C .

Butyl acrylate was obtained partly from I.C.I. (Dyestuffs Division), partly prepared in our laboratory from methyl acrylate (Lights Ltd.). The purification was

done in the same way as in the case of styrene, except that drying was effected with anhydrous sodium sulphate instead of with calcium chloride, as the latter forms complex compounds with butyl acrylate. Preparation of butyl acrylate from methyl acrylate was carried out by ester interchange with *p*-toluene sulphonic acid as a catalyst according to the recipe given in *Organic synthesis*, 26, 18. The butyl acrylate prepared in this way showed the same polymerization characteristics as the I.C.I. sample (same rate at the same benzoyl peroxide concentration at 25°C, no induction period; intensity exponent in photochemical polymerization 0.5). Benzoyl-peroxide was purified by dissolving it in chloroform, separating the water layer, filtration of the chloroform solution and precipitation in methanol. It was dried and stored in the dark *in vacuo* (20 mm.) over calcium chloride. Solutions of known concentration were made up in Analar benzene immediately before use, and a known volume of this was introduced into the dilatometers. The benzene was then evaporated *in vacuo*.

DETERMINATION OF MONOMER REACTIVITY RATIOS

The determination of the r_1 and r_2 values takes place on the basis of the copolymer composition equation of which the differential form is

$$\frac{d(A)}{d(B)} = \frac{(A)r_1(A) + (B)}{(B)r_2(B) + (A)}, \quad (1)$$

where $d(A)/d(B)$ is the ratio of A and B molecules adding to the radicals present in a copolymerizing system in which (A) and (B) are the concentrations of the monomers. If the composition of the monomer mixture remains virtually constant, which can be realized by working with a small percentage of conversion, $d(A)/d(B)$ can be taken as the composition of the copolymer formed. In this way the difficulty of working with the rather complicated integrated form of the copolymer composition equation is eliminated, and the experimentally difficult task of determining the ratio of (A) to (B) after a considerable conversion has taken place is avoided.

Theoretically, it would be sufficient to prepare copolymers from two starting mixtures of monomers of different composition. However, copolymers were made from five different starting mixtures to allow a satisfactory elimination of experimental errors and to get an idea about the accuracy of the results and the validity of the theory.

Though for a determination of r_1 and r_2 a very high degree of purity of the monomers is not absolutely required (Mayo, Lewis & Walling 1947), in our case monomers highly purified as for kinetic measurements were used, and the copolymers were prepared in dilatometers filled in a high vacuum. This procedure allowed the copolymerization to be followed adequately. The dilatometers were filled by immersion in liquid nitrogen and the monomers were distilled into them from burette tubes used as reservoirs. The amount of the monomers used was known within limits of ± 0.05 ml. Benzoyl peroxide was used as an activator. Copolymerization was effected at 25.0°C. Because the reaction proceeds very slowly, ultra-violet light of a mercury lamp was used to speed it up.

After about 10 % conversion the polymer was precipitated by pouring into excess of methanol at 0° C. The precipitate was filtered off, redissolved in a small amount of benzene and again precipitated in the same way. After intensive washing with methanol it was dried till constant weight at 100° C at atmospheric pressure, and subsequently it was freed from remaining amounts of solvent and precipitant by heating for short times in a molecular still (Grassie & Melville 1950) at a temperature of about 160° C. At this temperature the copolymers show the first sign of flow. After a first period of 5 min. at these conditions a loss in weight of several units per cent was found (solvent), but a second treatment did not cause any further change. The copolymers obtained in this way were used for analysis. The most elegant way to do this appeared to be a direct oxygen determination, the oxygen content being directly proportional to the butyl acrylate content. The method proposed by Schutze (1939), Unterzaucher (1940) and recently described by Degens & Gouverneur (1950) has been used.

The sample is pyrolyzed in a stream of nitrogen and the gases are then passed over carbon at 1100° C, where the oxygen present is quantitatively converted into carbon monoxide. The amount of this gas is found by allowing it to react with iodine pentoxide and determining either the amount of iodine or of carbon dioxide formed.

That this method may safely be applied to polymers can be deduced from an analysis of a sample of polybutylacrylate for which an oxygen content of 24.8 % (theory 25.0 %) was found. The experimental data are given in table 1.

TABLE 1

composition reaction mixture		composition reaction product	
butyl acrylate (ml.)	styrene (ml.)	oxygen (%)	weight fraction of butyl acrylate
7.7	1.9	15.7	0.628
6.5	2.8	14.0	0.560
4.9	4.9	11.7	0.468
3.5	7.0	8.9	0.356
2.4	8.5	7.5	0.300
Y	Z		X

As the density of butyl acrylate and of styrene is practically the same at the temperature at which they are distilled into the dilatometers (0.90 at 25 to 28° C), the volume ratios may be taken as weight ratios. As the unit in which concentration is expressed does not matter for the copolymer composition equation we can use Y and Z (see table 1) divided by the appropriate molecular weight of the monomer.

It is convenient for statistical analysis to use a modified form of the copolymer equation. The copolymer composition may be written as follows:

$$\frac{X}{128} \cdot \frac{104}{1-X} = 0.813 \frac{X}{1-X} = \frac{AX}{1-X}. \quad (2)$$

The copolymer composition equation is transformed into

$$r_2 X Z^2 + r_1 A (X - 1) Y^2 + (AX + X - 1) YZ = 0, \quad (3)$$

in which expression r_1 refers to the butyl acrylate radical and r_2 to the styrene radical.

The standard deviations of Z and Y , the volume readings were taken as 0.01667; the standard deviation of the oxygen determination as derived from a great number of experiments with benzoic acid amounts to 0.002667.

The values of r_1 and r_2 were calculated from the experimental data by the method of least squares, i.e. by minimizing the variance of the expression (3).

As the standard deviations of the experimental data are known the variance of (3) and the standard deviations of r_1 and r_2 follow by well-known methods. These standard deviations are a measure for the fluctuations in the values of r_1 and r_2 which would be found if the experiments were repeated. A second estimate of the variance of (3) is obtained by squaring the deviation of (3) from zero for every set of experimental values, adding and dividing by the difference between the number of experiments and the number of undetermined constants.

This last estimate gives an indication of the goodness of fit of the experimental points to the theory. If it is appreciably larger than the estimate of the variance of (3) based on the standard deviations of the experiments it should be concluded that the experiments allow of a better theoretical interpretation. However, both estimates gave nearly the same results, and consequently it may be concluded that the theory fits the experiments within experimental error.

The results of these calculations were:

$$r_1 = 0.16, \quad s_{r_1} = 0.04,$$

$$r_2 = 0.49, \quad s_{r_2} = 0.04.$$

In the calculations of which the results have been given the initial monomer composition of the mixture has been used. However, the composition at the end of the reaction (10 % conversion) might equally have been taken as representative for the monomer mixture. Therefore the values of r_1 and r_2 have also been calculated using instead of Y and Z in formula (3) Y' and Z' calculated from

$$Y' = Y - 0.1(Y + Z)X,$$

$$Z' = Z - 0.1(Y + Z)(1 - X).$$

In the same way as above it is found that

$$r_1 = 0.47, \quad s_{r_1} = 0.04,$$

$$r_2 = 0.13, \quad s_{r_2} = 0.04.$$

It follows that there is no significant difference between the two sets of r_1 and r_2 values. Consequently, the use of the copolymer composition equation in its form (1) is justified.

Further calculations will be based on the average of the two r_1 and r_2 values

$$r_1 = 0.48, \quad s_{r_1} = 0.04,$$

$$r_2 = 0.15, \quad s_{r_2} = 0.04.$$

Figure 1 gives the relationship between percentage butyl acrylate in copolymer and in monomer mixture and shows that the 'azeotropic' mixture is between 40 and 45 % of butyl acrylate. At lower percentages butyl acrylate is preferably incorporated

into the copolymer, at higher percentages, however, there is a preference for styrene. A rather strong tendency for alteration is shown by the product $r_1 r_2$, which is rather low (0.072). It is possible to calculate the Alfrey-Price Q and e factors for butyl acrylate taking these factors respectively 1 and -1 for styrene. One arrives at a Q value of 0.4 and an e value of 0.6. The low Q value compared with that of styrene points to less resonance possibilities in the butyl acrylate radical than in the styrene one; the value of e is positive, and therefore the double bond has a relatively positive character which may be an important factor in explaining the preference of the

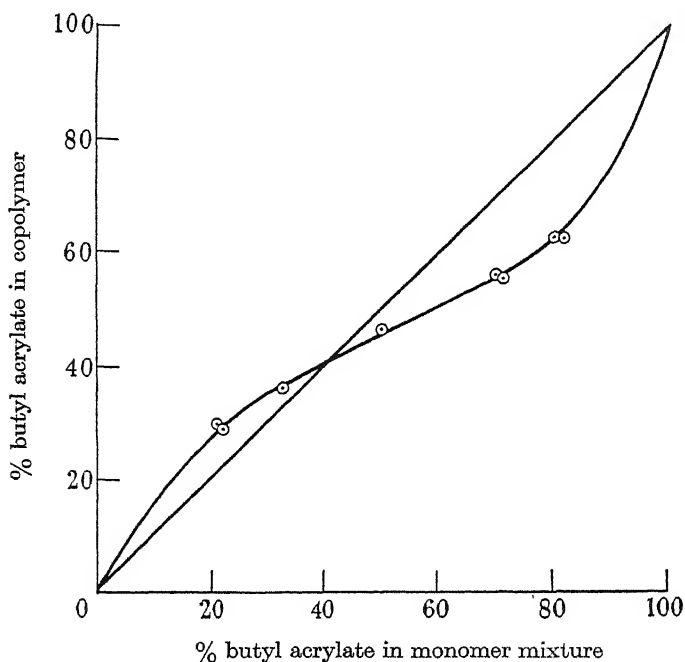


FIGURE 1. The relation between monomer mixture and copolymer composition.

radicals for unlike monomers or radicals which is found in this system. Q and e values for methyl acrylate $Q = 0.35$, $e = 0.5$ given by Price (1947) compare rather well with the values given above, as the length of the carbon chain in the ester group of the acrylate can be expected not to have a very pronounced influence on the general reactivity of the radical or the polarity of the double bond. In fact, for the systems vinylidene chloride-ethyl methacrylate and vinylidene chloride-butyl methacrylate (Alfrey and co-workers 1948) the same r_1 and r_2 values have been found.

However, in view of the inaccuracy in various r_1 and r_2 values and the different temperatures at which they have been determined (see the survey of available data given by Simha & Wall 1948), and, further, the semi-empirical character of the Alfrey-Price relations, further speculations on differences of Q and e values probably are not justified.

It will be shown in part II that the accuracy in the r_1 and r_2 values is great enough to allow a reliable analysis of the kinetics of the copolymerization of butyl acrylate and styrene.

II. THE INITIATION AND TERMINATION MECHANISM

INTRODUCTION

In part I the determination of the monomer reactivity ratios for the system styrene-butyl acrylate was described. The knowledge of these ratios makes it possible to evaluate the kinetic coefficients of initiation and termination from measurements of rate of copolymerization and of osmotic molecular weights of the copolymers formed.

METHODS USED

Copolymerization was investigated in mixtures of styrene and butyl acrylate at 25°C. As a catalyst only benzoyl peroxide was used, which was dissociated by ultra-violet radiation from a mercury lamp. The dilatometers were made of Pyrex glass, which absorbs wave-lengths less than 2700 Å. Filling of the dilatomers was done in absence of oxygen ($< 10^{-4}$ mm.). The monomers were degassed thoroughly by repeated freezing and melting in their reservoirs which were connected with the vacuum line and then distilled into graduated tubes. From these a known amount of monomers could be distilled at a temperature of about 25°C into the dilatomers immersed in liquid air. Eventually, traces of air entrapped in the benzoyl peroxide were removed by a last degassing carried out in the dilatometer. After sealing off, the dilatometers were stored in liquid air till used. Immediately before use the contents were thoroughly mixed by shaking.

Rates of copolymerization were determined at a temperature of 25.0°C, the temperature being kept constant within 0.01°C. In those experiments in which the dilatomers were irradiated an extra cooling was provided by a coil through which tap water flowed under a constant pressure head.

The height of the meniscus in the stem of the dilatometer was measured with the aid of a cathetometer reading to 0.001 cm. The dilatomers were calibrated with mercury, uniformity of the section of the stem over its whole length having been ascertained.

The exact relation between the volume contraction measured and the extent of copolymerization of both monomers is not known; theoretically, it is necessary to take into account not only the composition of the copolymer being formed, but also the composition of the monomer mixture. However, in this case where (a) the monomers mix without volume contraction, (b) copolymer composition does not deviate very strongly from that of the monomer mixture, and (c) contractions for 100 % polymerization for the separate monomers do not differ considerably (13.4 % for butyl acrylate and 14.4 % for styrene), it is unimportant whether interpolation is made between the contractions mentioned on the basis of copolymer or monomer mixture composition as was done here. The difference is small compared with experimental errors arising from other sources.

The low-pressure mercury lamp was operated at a voltage of 1000 V from a transformer fed through a constant voltage transformer. By an adjustable resistance the current could be kept constant at 3 ± 0.1 A. The lamp was not used until half an hour after switching on.

The transmission of Pyrex glass for ultra-violet radiation decreases rapidly below 3000 Å, the wave-length below which styrene and benzoyl peroxide show a strong absorption. On the other hand, the light emitted by the low-pressure mercury lamp practically is confined to wave-lengths above 3000 Å. Therefore it may be assumed that intensity of radiation through the reaction tubes has been nearly uniform and consequently also concentration of radicals. This assumption is supported by experiments of Burnett, Valentine & Melville (1949), in which it was proved that in the ultra-violet activated (2950 Å) polymerization of styrene with benzoyl peroxide as a catalyst uniformity of radical concentration exists. Butyl acrylate shows only a strong absorption further in the ultra-violet (see appendix 1).

For the determination of molecular weights, copolymers were isolated in the way described in part I. Osmotic molecular weights were determined at 25.0° C in dry benzene as a solvent. The osmometer used was of the Fuoss-Mead type modified as described by Masson & Melville (1949).

EXPERIMENTAL DATA

Some experiments were carried out to find whether in this system termination takes place by mutual interaction of growing radicals as is the case with both monomers. The rate of copolymerization should be then proportional to the square root of benzoyl peroxide concentration and of the light intensity.

The experiments were carried out with a butyl acrylate-styrene mixture of about the 'azeotropic' composition, which has the advantage that composition of the monomer mixture does not change with amount of conversion. Consequently as long as no gel effect interferes, the same tube may be used under different conditions. The data are given in table 2 and in figures 2 and 3.

TABLE 2. DATA IN RELATION TO TERMINATION MECHANISMS

weight ratio butyl acrylate styrene	conc. benzoyl peroxide (10 ⁻³ g. mol./l.)	rate normal no u v. (% per hr.)	rate u.v. full intensity (% per hr.)	rate u.v. with screen* (% per hr.)	intensity exponent
45 : 55	{ 125.6	0.27	2.11	—	—
	{ 47.5	0.20	1.32	0.74	0.50
	{ 26.2	0.14	1.89†	1.095†	0.46
	{ 0	0	0.21†	0.12†	0.47

* A perforated screen with a transmission of 31 % was used.

† In these experiments a more powerful ultra-violet lamp than usual was used.

It will be seen that the experimental facts are in agreement with a mutual termination mechanism.

Consequently for evaluation of the kinetic data the following expression (Melville *et al.* 1947) may be employed:

where

$$-\frac{d(A+B)}{dt} = \text{rate of copolymerization,}$$

(A) and (B) = concentrations of butyl acrylate and styrene respectively,

r_1 and r_2 = reactivity ratios for butyl acrylate and styrene respectively,

δ_a and $\delta_b = \frac{k_t^\dagger}{k_p}$ for butyl acrylate and styrene respectively,

X = rate of initiation, and

$$\phi = \frac{k_{iab}}{(k_{iaa}k_{ibb})^{\frac{1}{2}}}.$$

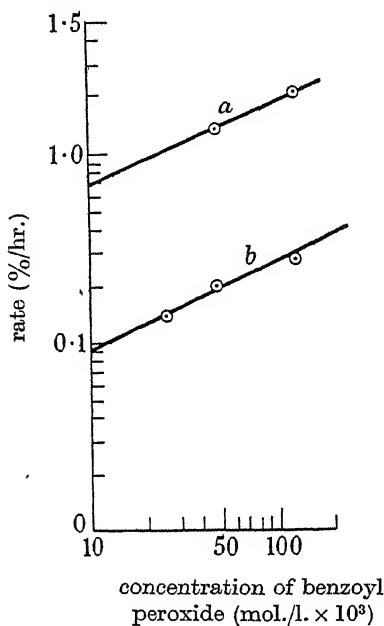


FIGURE 2

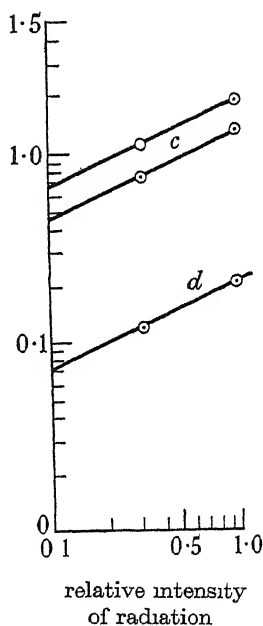


FIGURE 3

FIGURES 2, 3. The variation of rate with catalyst concentration (figure 2) and with light intensity (figure 3). *a*, with; *b*, without ultra-violet radiation; *c*, with; *d*, without catalyst.

The experimental data obtained are given in table 3.

Some comment is necessary in regard to these data. First, the thermal rate of copolymerization without using ultra-violet activation was measured. Subsequently the dilatomers were exposed to the radiation of the ultra-violet lamp. The contributions to conversion in both conditions were determined, since this is necessary for calculation of the average kinetic chain length of the resulting copolymers, which consist of a part made under thermal conditions (high molecular weight) and a part made under ultra-violet activated conditions (low molecular weight).

In some cases (copolymer 6 and G) some changes in position of lamp and dilatometer occurred during the run; consequently two ultra-violet rates and percentages of conversion are given in table 2.

A better insight of the kinetic features of this system is obtained by calculating the rates which would have been obtained if all experiments were carried out at the same concentration of benzoyl peroxide. A concentration of 40×10^{-3} g.mol./l. has been chosen as reference standard, since for this concentration data already existed for polystyrene (Melville & Valentine 1950*a*).

TABLE 3. EXPERIMENTAL DATA. RATE AND MOLECULAR WEIGHT MEASUREMENTS

copolymer	composition (% butyl acrylate)		benzoyl peroxide conc. (mg.mol./l.)	rate $\times 10^5$ (g.mol./l./sec.)		% thermal	number
	monomer mixture	co- polymer		thermal	u.v.	conversion	average mol.wt.
						% u.v. conversion	M_n
E	96	90	53.7	2.2	30.2	1.10/11.32	140,000 \pm 20,000
C	90	78	45.3	0.62	5.66	0.85/5.25	125,000 \pm 10,000
6	80	63	12.2	0.24	0.45	1.5/0.4	530,000 \pm 100,000
					0.88	1.5/8.0	
3	50	47	36.6	0.31	2.88	0.7/9.1	87,000 \pm 5,000
G	46	45	62.5	0.44	3.92	0.7/2.38	63,000 \pm 5,000
					3.55	0.7/3.33	
B	31	36	58.3	0.38	3.44	0.55/6.12	64,000 \pm 4,000
5	22	30	56.7	0.28	2.12	0.25/11.0	77,000 \pm 10,000
D	10	18	56.7	0.37	3.17	0.51/5.26	59,000 \pm 5,000
butyl acrylate	100	100	40	26.9	—	—	—
styrene	0	0	40	0.48	—	—	—

Figure 4 gives the curve calculated for the rate of the thermal copolymerization at this reference benzoyl peroxide concentration as a function of composition of monomer mixture. It is seen that the addition of a few units per cent of styrene to butyl acrylate causes a large drop in rate, a minimum being reached apparently at about 80 % of styrene.

Of the three factors in the equation for the rate the first term in the numerator of (4) cannot possibly account for the drop in rate. This term varies relatively slightly with composition of the reaction mixture and, moreover, in the wrong direction, i.e. increasing with increasing styrene content.

Therefore either the initiation term in the numerator or the termination term in the denominator including ϕ or both must be responsible for the form of the rate-composition curve.

The results will be discussed on the basis of the assumption that the initiation factor may be written as

$$X = C_{b.p.o.}[k_{ia}(A) + k_{ib}(B)] \quad (5)$$

in which expression k_{ia} is the reaction constant of the initiation reaction between butyl acrylate and the radical derived from benzoyl peroxide, and k_{ib} the corresponding one for styrene, which both should be derived from data on the benzoyl peroxide induced polymerization of the separate monomers. This may be done with the aid of the equation

$$-\frac{d(M)}{dt} = \frac{C_{b.p.o.}^{\frac{1}{2}} k_i^{\frac{1}{2}}(M)^{\frac{3}{2}}}{\delta} \quad (6)$$

For butyl acrylate therefore ($\delta = 7.6$)

$$k_{ia} = 2.05 \times 10^{-7}.$$

For styrene ($\delta = 71$)

$$k_{ib} = 4.51 \times 10^{-9}.$$

Consequently $X = C_{b.p.o.}(2.05 \times 10^{-7}(A) + 4.51 \times 10^{-9}(B))$.

With the value for X calculated in this way the values for ϕ may be derived immediately from rate experiments.

It then turns out that the factor ϕ should have a very high value. Only from the rate in the mixture containing but 4 % styrene, a ϕ factor smaller than 1000 is computed (140), but the results for all other mixtures give ϕ values ranging from 10^3 to 4×10^3 . This would mean that termination by unlike radicals is very strongly preferred to termination by like radicals. However, if on the basis of these high ϕ values the average degree of polymerization of the copolymers is calculated, the values are so low that it would be hardly possible to measure the average molecular weight of the copolymers osmotically. As this was not at all the case it appears that rate of initiation in this copolymer system is not a linear function of composition and the initiation constants for the separate monomers. Consequently, an experimental determination of the rate of initiation becomes necessary.

The osmotic pressure measurements of solutions of the copolymers in benzene are given in table 4.

These results are not quite accurate enough to state the existence or absence of a trend in Huggins μ value with composition of copolymer as found by Melville & Valentine (1950b) for copolymers of methyl methacrylate and styrene. It may only be stated here that the Huggins μ value for copolymers of butyl acrylate and styrene is in the region 0.50 to 0.40. For the calculation of the trend in rate of initiation with composition of the monomer mixture and the estimation of the factor ϕ , the average kinetic chain length has been calculated assuming that the mechanism of mutual termination is a disproportionation reaction. To this end it is necessary to take into account that the copolymers really prepared consisted of at least two parts as mentioned in table 3.

Let

n_n = weight fraction of total copolymers made under thermal conditions,

n_u = weight fraction of total copolymers made under ultra-violet conditions,

R_n = rate during thermal polymerization (no ultra-violet),

R_u = rate during ultra-violet polymerization,

X_n = rate of initiation during thermal reaction,

X_u = rate of initiation during ultra-violet activation,

\bar{m} = the average molecular weight of the monomer in the copolymer to be calculated from its composition,

\bar{M} = osmotic molecular weight of the copolymer,

ν_t = average degree of polymerization of the copolymer,

ν_n = average degree of polymerization of the thermal copolymer,

ν_u = average degree of polymerization of the ultra-violet copolymer,

thus the rate of initiation may be calculated as follows:

For the thermal and the photo-activated reaction the following relations hold:

$$\bar{\nu}_n = \frac{R_n}{X_n}, \quad \bar{\nu}_u = \frac{R_u}{X_u}.$$

For X_u may be substituted $\left(\frac{R_u}{R_n}\right)^2 X_n$,

and as $\frac{1}{\bar{\nu}_t} = n_n \frac{1}{\bar{\nu}_n} + n_u \frac{1}{\bar{\nu}_u}$,

then $\frac{1}{\bar{\nu}_t} = \frac{1}{\bar{P}} = \frac{\bar{m}}{\bar{M}} = X_n \left(\frac{n_n}{R_n} + n_u \left(\frac{R_u}{R_n^2} \right) \right).$ (7)

As the rate of initiation may be taken as proportional to concentration, the rate of initiation may be calculated at the reference concentration of 40×10^{-3} g. mol./l. of benzoyl peroxide.

The results obtained are given in table 5 and in figure 5.

TABLE 4. OSMOTIC DATA

co-polymer	com-position (% butyl acrylate)	conc. solution (g./100 g. of solution) (c)	osmotic pressure (cm. of solution) (P)	P/c	lim P/c $c \rightarrow 0$	\bar{M}_n
E	90	0.890	2.36	2.67	1.8	140,000 \pm 20,000
		0.574	1.46	2.54		
		0.314	0.63	2.00		
C	78	0.931	2.79	3.00	2.0	125,000 \pm 10,000
		0.547	1.40	2.56		
		0.261	0.64	2.45		
6	63	1.36	3.79	2.78	0.47	530,000 \pm 100,000
		0.885	1.63	1.84		
		0.548	0.78	1.42		
3	47	0.514	2.24	4.36	2.9	87,000 \pm 5,000
		0.337	1.27	3.78		
		0.182	0.64	3.52		
		0.113	0.34	3.10		
G	45	0.932	3.70	3.97	4.0	63,000 \pm 5,000
		0.444	1.71	3.86		
		0.250	1.05	4.20		
		0.156	0.63	4.04		
B	36	0.904	4.00	4.42	3.95	64,000 \pm 4,000
		0.593	2.51	4.24		
		0.342	1.41	4.13		
5	30	1.35	7.92	5.86	3.3	77,000 \pm 10,000
		0.780	3.42	4.38		
		0.627	2.52	4.02		
		0.359	1.46	4.07		
		0.296	1.17	3.96		
D	18	0.809	3.75	4.64	4.2	60,000 \pm 5,000
		0.670	3.265	4.86		
		0.160	0.69	4.32		

TABLE 5

copolymer	overall rate g. mol./l./sec. $\times 10^5$	rate of initiation $\times 10^{10}$	ϕ
E	1.9	15.7-11.8	2.5-1.4
C	0.59	7.8-6.7	14-12
6	0.41	4.7-3.2	20-11
3	0.32	5.2-4.6	39-27
G	0.35	9.0-7.6	91-73
B	0.31	7.0-6.2	97-82
5	0.24	5.5-4.3	151-93
D	0.31	7.8-6.6	143-95
butyl acrylate	26.9	580	
styrene	0.48	15.5	

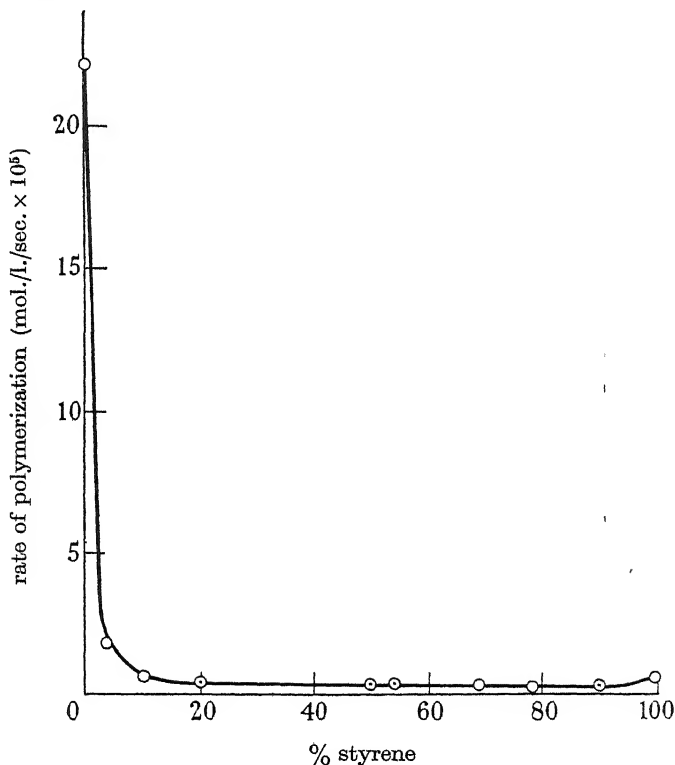


FIGURE 4. Rate of copolymerization at 25° C as a function of the composition of the monomer mixture ($C_{b.p.o.} = 4 \times 10^{-2}$ mol./l.).

From table 5 it is clear that the kinetics of the copolymerization of styrene with butyl acrylate require the introduction of a factor describing the rate of termination between unlike radicals. This factor is rather high, of the order of magnitude of 10^2 , which, however, is 10 times smaller than was calculated from the previously mentioned assumption⁷ as regards rate of initiation in a copolymerizing system. As is seen in table 5 no constant ϕ value is found in this case.

However, it has to be kept in mind that the unavoidable experimental errors, and the uncertainties in the reaction constants for both monomers are all reflected in the value of ϕ .

It is found by calculation that even if the rather exaggerated assumption is made that the true r_1 and r_2 values differ from the determined ones by a factor 2, this has not a very pronounced influence on ϕ .

As (A) and (B) are known with accuracy it is not to be expected that large variations in ϕ are due to errors in these quantities except at very low (A) or (B). Perhaps the low value of ϕ for the case of copolymer E may be partly a consequence of this circumstance but a factor of more than 2 seems quite improbable.

As regards the measurements of the rate of copolymerization errors of more than 10 % need not be assumed, and the same applies to measurements of osmotic molecular weights.

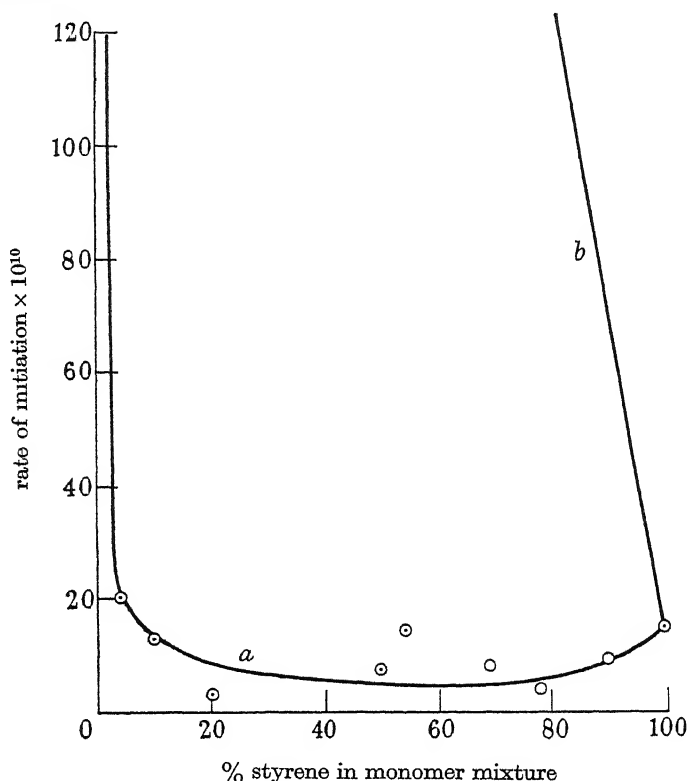


FIGURE 5. Rate of initiation as a function of monomer composition ($C_{b.p.o} = 4 \times 10^{-2}$ mol./l.).
a, experimental; b, calculated from rates of initiation of the separate monomers.

The uncertainties in the value δ_a and δ_b naturally are very important for the value of ϕ . Too high values of δ_a and δ_b lead to a lower ϕ .

The theory of copolymerization as at present formulated assumes that ϕ should be constant and this has been found for the styrene-methyl methacrylate system. The results obtained here vary more than could be possibly explained by experimental error. It may therefore be necessary to modify even still further copolymerization theory to take into account such a new effect. It may be, for example, that environment may have an effect in termination velocity coefficients.

The complete set of reaction constants for the system butyl acrylate-styrene at 25°C is given in table 6. The reaction constants for styrene are those given by

Melville & Valentine (1950b), those for butyl acrylate were determined by Bickel & Melville (1949).

The values of the propagation constant for butyl acrylate have been determined by Bickel & Melville by two different methods.

Though the agreement between the two values is satisfactory for this kind of work, it naturally entails the same inaccuracy in the reaction constant of the reaction of the butyl acrylate radical with the styrene molecule. The calculations of ϕ have been carried out with the average value of

$$\delta_a = k_{iaa}^\dagger / k_{paa}$$

being 7.6 (average of 5.6 and 9.6).

TABLE 6

reaction	symbol	numerical value
butyl acrylate radical-molecule	k_{paa}	22.5
		13.1
butyl acrylate radical-radical	k_{iaa}	1.6×10^4
styrene radical-molecule	k_{pbb}	39.5
styrene radical-radical	k_{ibb}	7.9×10^4
butyl acrylate radical-styrene molecule	k_{pab}	150
		87.5
styrene radical-butyl acrylate molecule	k_{pba}	82.4
butyl acrylate radical-styrene radical	k_{iab}	$c. 5 \times 10^7$

From the results given in table 6 it is seen that the butyl acrylate radical preferably reacts with a styrene molecule and a styrene radical, whereas the styrene radical preferably reacts with a butyl acrylate molecule and a butyl acrylate radical. In part III, it will be shown that this preference for 'crossed' reaction probably may be linked up with polar effects playing a role as well in the reactions of radicals with monomers as in the reactions of radicals with each other.

III. DISCUSSION OF RESULTS

The experiments leading to the evaluation of the rate constants in the copolymerization system butyl acrylate-styrene have been described. However, it remains to be discussed whether the assumptions made are valid and to try to give an explanation of the particular features of the system investigated.

A first assumption which has been made is the absence of transfer reactions. The occurrence of transfer theoretically is of importance for two reasons:

(1) It has been shown by Melville *et al.* (1947) that the r_1 and r_2 values determined for a copolymerizing system with the aid of the copolymer composition equation are not the true quotients of the reaction constants of a radical with the two monomers present but the products of these quotients with a certain factor dependent on transfer; thus

$$r'_1 = r_1 \frac{(1 + \epsilon_{aa})}{(1 + \epsilon_{ab})}, \quad r'_2 = r_2 \frac{(1 + \epsilon_{bb})}{(1 + \epsilon_{ba})},$$

where

$$\epsilon_{aa} = \frac{k_{faa}}{k_{paa}}, \quad \epsilon_{ab} = \frac{k_{fab}}{k_{pab}}, \quad \text{etc.}$$

The transfer ratios of butyl acrylate and styrene, however, are both of the orders of magnitude of 10^{-5} . Even if the crossed transfer constants were much higher the influence of transfer on the r_1 and r_2 values found would fall within the experimental error in their determination given in part I.

(2) The most important influence of transfer would be that on rate of initiation, as this is determined from molecular weight data.

For homopolymerization in bulk with transfer taking place the expression holds

$$\frac{1}{\bar{P}} = \frac{1}{\bar{\nu}} + \epsilon_{aa}; \quad (8)$$

for the case of copolymerization we obtain

$$\frac{1}{\bar{P}} = \frac{1}{\bar{\nu}} + \Sigma \epsilon_{ab}, \quad (9)$$

where P is the degree of polymerization, ν is the kinetic chain length and $\Sigma \epsilon_{ab}$ stands for the quotient of the rate of transfer by all processes and the rate of propagation by all processes.

It seems improbable that the crossed transfer ratios are much higher than the corresponding ones for the monomers.

However, the same factors which are responsible for the deviations of r_1 and r_2 from unity and for the high ϕ value might enhance the transfer ratios. Transfer ratios for the separate monomers being of the order of magnitude 10^{-5} (Melville & Valentine 1950*b*), it is felt that the transfer ratio to be taken into account could be perhaps of the order of magnitude 10^{-4} . For molecular weights well below 100,000, i.e. $\bar{P} < 10^3$, a transfer ratio of 10^{-4} will cause only an error of 5 to 10 % in X . However, for molecular weights well above 100,000 the neglect of transfer will be more serious. We therefore have recalculated X and consequently ϕ for those cases on the assumption of $\Sigma \epsilon_{ab} = 10^{-4}$.

We then find for the case of the copolymers with a molecular weight higher than 100,000 the data, given in table 7. As was to be expected the influence of transfer is most pronounced with copolymer 6 of a molecular weight of about 5×10^5 .

TABLE 7

copolymer	rate of initiation $\times 10^{10}$		ϕ	
	without transfer	with transfer	without transfer	with transfer
E	c. 14	c. 12	2	1.5
C	c. 7	c. 6.5	13	10
6	c. 4	c. 2.2	15	7

Even in these cases the influence of the assumed transfer ratio on the value of ϕ is not such that ϕ is brought down to about 1. An improbably high transfer ratio would have to be assumed to account by transfer for the peculiarities of the system investigated.

A second assumption that has been made throughout is that the mutual termination process is by disproportionation and not by combination.

In the case of homopolymerization experimental evidence—though not very abundant—is generally in favour of a disproportionation mechanism. However in the case of reaction between unlike radicals where polar factors of opposite sign might play a role, termination by combination might occur.

Assuming this to be the case allowance has to be made for the fact that for disproportionation

$$\bar{P} = \frac{R}{\bar{X}}, \bar{P} = \text{average degree of polymerization,}$$

whereas for combination
$$\bar{P} = \frac{2R}{\bar{X}}.$$

If a fraction x of the radicals combine then

$$X = \frac{R}{\bar{P}(1-0.5x)}. \quad (10)$$

Now x can be expressed in terms of the reaction constants and concentration of radicals if it is assumed that unlike radicals combine

$$x = \frac{k_{iab}(P)(Q)}{k_{iaa}(P)^2 + k_{iab}(P)(Q) + k_{ibb}(Q)^2},$$

which amounts to

$$x = \frac{2r_1 r_2 \delta_a \delta_b (A)(B)\phi}{\delta_a^2 r_1^2 (A)^2 + \delta_b^2 r_2^2 (B)^2 + 2r_1 r_2 \delta_a \delta_b (A)(B)\phi}, \quad (11)$$

or
$$\frac{1}{x} = 1 + \frac{r_1^2 \delta_a^2 (A)^2 + r_2^2 \delta_b^2 (B)^2}{2r_1 r_2 \delta_a \delta_b (A)(B)\phi} = 1 + \frac{Q}{\phi}. \quad (12)$$

In table 8 the values of Q for the monomer mixtures investigated are given together with the values of x and the correction factor with which the rate of initiation calculated before has to be multiplied if unlike radicals combine.

TABLE 8. INFLUENCE OF ASSUMED COMBINATION OF UNLIKE RADICALS ON RATE OF INITIATION FOR DIFFERENT ϕ 's. VALUES FOR x AND CORRECTION FACTOR (c.f.)

composition monomer mixture (% butyl acrylate)	Q	$\phi = 1$		$\phi = 10$		$\phi = 100$		$\phi = 1000$	
		x	c.f.	x	c.f.	x	c.f.	x	c.f.
96	1.04	0.49	1.32	0.91	1.84	0.99	1.98	1.0	2
90	2.21	0.31	1.18	0.82	1.70	0.98	1.96	1.0	2
80	4.72	0.175	1.10	0.68	1.52	0.96	1.92	0.995	2
50	18.5	0.05	1.03	0.35	1.21	0.84	1.72	0.98	1.96
46	21.7	0.045	1.02	0.32	1.19	0.82	1.70	0.98	1.96
31	41.1	0.025	1.01	0.20	1.11	0.71	1.55	0.96	1.92
22	65.2	0.02	1.01	0.13	1.07	0.61	1.44	0.94	1.89
10	167	0.006	1.00	0.05	1.03	0.36	1.22	0.86	1.76

As ϕ appears to be of the order of magnitude of 10^2 , the corresponding correction factor has been applied to see whether a ϕ -value of better constancy is obtained. A comparison of rate of initiation and ϕ calculated on this basis is given in table 9.

TABLE 9. INFLUENCE OF ASSUMED COMBINATION OF UNLIKE RADICALS ON RATE OF INITIATION AND ϕ

copolymer	rate of initiation X		ϕ	
	disproportionation	partial combination	disproportionation	partial combination
E	15.7-11.8	31.4-23.6	2.5-1.4	7.2-5.2
C	7.8-6.7	15.6-13.4	14-12	30-25
6	4.7-3.2	8.9-6.1	20-11	42-28
3	5.2-4.6	8.4-7.8	39-27	94-58
G	9.0-7.6	15.3-12.9	91-73	170-140
B	7.0-6.2	10.9-9.6	97-82	174-149
5	5.5-4.3	7.9-6.2	151-93	248-180
D	7.8-6.6	9.5-8.0	143-95	208-150

It is seen that in this way there is some effect in the direction of flattening the ϕ -composition curve and smoothing the X -composition curve, but even so a constant value for the whole range of composition is not obtained.

Consequently, we come to the conclusion that neither the phenomenon of transfer nor the possibility of combination of unlike radicals allow of a satisfactory explanation for the particular features of the system butyl acrylate-styrene.

It is therefore presumed that in the initiation process the explanation for the anomalies has to be found. It is very strange the rate of initiation deviates so much from linearity with composition. It is necessary to bear in mind that the calculation of rate of initiation has been based on results of osmotic molecular weight determinations of copolymers isolated. The possibility is not excluded that the rate of decomposition of benzoyl peroxide was linear or nearly linear with composition but that the radicals formed by addition of say one monomer molecule to the radicals of the catalyst react so quickly with each other that compounds of low molecular weight are formed which escape precipitation. Osmotically only the 'large' polymer molecules are detected and therefore the rate of initiation calculated in this way would in fact be abnormally low.

A first purely mathematical approach towards a kinetical description of this kind of process would be to write rate of initiation—as far as it leads to truly polymeric molecules—as

$$X = C\{k_{ia}(A) + k_{ib}(B) - k_{ab}(A)(B)\}. \quad (13)$$

If k_{ab} is calculated from the experimental data then k_{ab} should be of the order of magnitude 10^{-7} , as is the case with the initiation constant for benzoyl peroxide-butyl acrylate. It varies with composition, being 5.3×10^{-7} for 96 %, 2.3×10^{-7} for 90 %, 0.9 for 80 % butyl acrylate, and for less butyl acrylate of the order of magnitude 0.5 to 0.3×10^{-7} .

Experimental proof for the formation of low molecular weight compounds cannot be given. The amount—if formed—would be so small as to be hardly detectable.

During the preparation of this paper the work of Walling (1949) on overall rates of copolymerization² has been published including the investigation of the styrene-methyl acrylate system. The most important difference with this work is the use of 2-azo-*bis*-isobutyronitrile as a catalyst instead of benzoyl peroxide. In his investigation it is assumed that rate of initiation is constant through the whole range of

composition, and in this way he arrives at a value for ϕ in the system styrene-methyl methacrylate which is in good agreement with that of Melville & Valentine. Though less pronounced than in the case of styrene-butyl acrylate a tendency for increase of ϕ with styrene content is present.

The system styrene-methacrylate closely resembles the system styrene-butyl acrylate as regards the shape of the rate-composition curve and the rather large variations of ϕ with small variations in rate

If the assumption of a constant rate of initiation at a certain temperature and at a certain concentration of azo catalyst through the whole range of composition of monomer mixtures is right, which still has to be proved experimentally, the suggestion put forward above would have to be rejected. Then the question how to explain the deviation of initiation from linearity with composition in systems with benzoyl peroxide as a catalyst still would remain open.

A question not yet discussed is that of finding an explanation for the high value of ϕ . It was mentioned above that styrene and butyl acrylate radicals probably have different electrical charges, styrene being negative, butyl acrylate positive, according to the system developed by Alfrey & Price (1947).

In a quite analogous way, as these authors did for reactions between monomers and radicals, one may write (Arlman, Melville & Valentine, 1949) the termination constants as a product of a probability factor, a general reactivity factor and an electrical charge factor of the radicals concerned. Then it is easily derived that

$$\phi^2 = \frac{A_{ab}^2}{A_{aa}A_{bb}} \frac{1}{r_1 r_2}.$$

Now $(r_1 r_2)^{-1}$ in this case is 14 and ϕ is about 100.

Substituting the values for A_{aa} and A_{bb} given by Bickel & Valentine, i.e. 1.8×10^4 for butyl acrylate and 2.2×10^8 for styrene, there is obtained

$$A_{ab} = 5.3 \times 10^7.$$

As the average value of k_{tab} was found to be 5.7×10^7 the energy of activation for the reaction between unlike radicals is practically zero.

The parallelism between ϕ and $(r_1 r_2)^{-1}$ may be demonstrated by table 10, in which only are mentioned those pairs of monomers from which a copolymer is formed soluble in the reaction mixture, in which case consequently there are no complications due to inhomogeneity.

TABLE 10. RELATION OF $\frac{1}{r_1 r_2}$ AND ϕ

monomer pair	$1/r_1 r_2$	ϕ
styrene-methyl methacrylate	4.2	17
styrene-methyl acrylate	6.6	40
styrene-butyl acrylate	14	100

Though certain details still remain to be explained, we feel that the investigation reported allows of the conclusion that the schemes now existing for the explanation of the phenomena in copolymerization seem to be adequate for a quantitative description.

One of us (E.J.A.) wishes to express his gratitude to the Management of the Koninklyke Shell-Laboratorium, Amsterdam for the opportunity to carry out this investigation. Thanks are also due for valuable help given by the theoretical analytical and spectroscopic departments of that Laboratory.

APPENDIX. THE ULTRA-VIOLET ABSORPTION SPECTRUM OF STYRENE,
BUTYL-ACRYLATE AND BENZOYL PEROXIDE

The ultra-violet absorption of both monomers were determined at Koninklyke Shell-Laboratorium at Amsterdam with the aid of a Beckmann quartz spectrophotometer model DU. Ethyl alcohol was used as a solvent; concentration of monomers was about 1.5×10^{-2} g.mol./l.; cell length 0.01 cm.

The data for styrene are in close agreement with previously reported ones (Rodebush & Feldman 1946, Smakula 1934).

The data for benzoyl peroxide are those given by Cohen, Ostberg, Sparrow & Blout (1948).

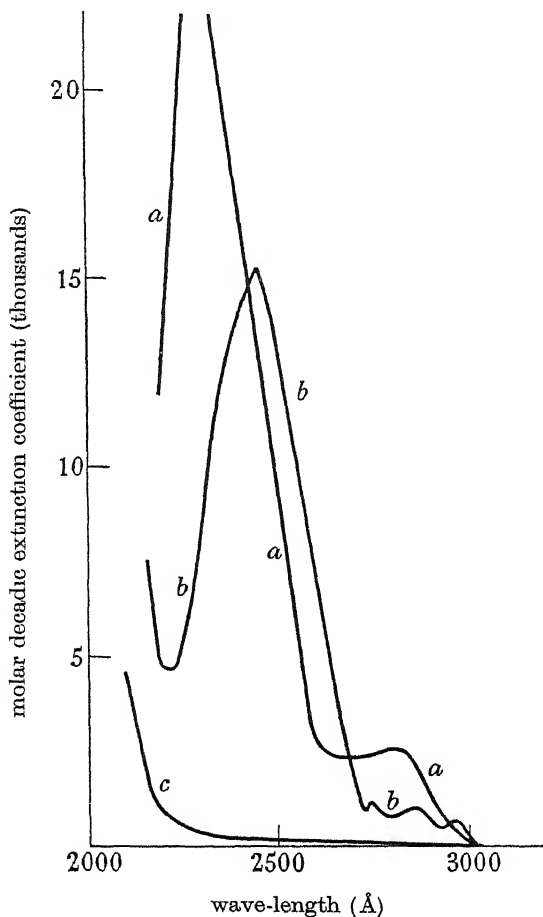


FIGURE 6. Absorption spectra of *a*, benzoyl peroxide; *b*, styrene, *c*, butyl acrylate in ethyl alcohol.

REFERENCES

- Alfrey, T., Agron, P., Bohner, J., Haas, H. & Wechsler, H. 1948 *J. Poly. Sci.* **3**, 157.
Alfrey, T. & Goldfinger, G. 1944 *J. Chem. Phys.* **12**, 205, 322.
Alfrey, T. & Price, C. C. 1947 *J. Poly. Sci.* **2**, 101.
Arlman, E. J., Melville, H. W. & Valentine, L. 1949 *Rec. Trav. chim. Pays-Bas*, **68**, 945.
Bickel, A. F. & Melville, H. W. 1949 *Trans. Faraday Soc.* **45**, 1049.
Burnett, G. M., Valentine, L. & Melville, H. W. 1949 *Trans. Faraday Soc.* **45**, 960.
Cohen, S. G., Ostberg, B. E., Sparrow, D. B. & Blout, E. R. 1948 *J. Poly. Sci.* **3**, 269.
Degens, P. N. & Gouverneur, P. 1950 *Anal. Chim. Act.* (in the Press).
Grassie, N. & Melville, H. W. 1950 *J. Poly. Sci.* (in the Press).
Masson, C. R. & Melville, H. W. 1949 *J. Poly. Sci.* **4**, 337.
Mayo, F. R. & Lewis, F. M. 1944 *J. Amer. Chem. Soc.* **66**, 1594.
Mayo, F. R., Lewis, F. M. & Walling, C. 1947 *Disc. Faraday Soc.* No. 2, 287
Melville, H. W., Noble, B. & Watson, W. F. 1947 *J. Poly. Sci.* **2**, 229.
Melville, H. W. & Valentine, L. 1950a *Proc. Roy. Soc. A*, **200**, 353.
Melville, H. W. & Valentine, L. 1950b *Trans. Faraday Soc.* **46**, 210.
Price, C. C. 1947 *Disc. Faraday Soc.* No. 2, 307.
Rodebush, W. H. & Feldman, I. 1946 *J. Amer. Chem. Soc.* **68**, 897.
Schutze, M. 1939 *Z. anal. Chem.* **118**, 241
Simha, R. & Wall, L. A. 1948 *Bur. Stand. J. Res., Wash.*, **41**, 521.
Smakula, A. 1934 *Z. angew. Chem.* **47**, 777.
Unterzaucher, J. 1940 *Ber. deutsch. Chem. Ges.* **73B**, 391.
Walling, C. 1949 *J. Amer. Chem. Soc.* **71**, 1930.
-

Polypeptide chain configurations in crystalline proteins

BY SIR LAWRENCE BRAGG, F.R.S., J. C. KENDREW AND M. F. PERUTZ

Cavendish Laboratory, University of Cambridge

(Received 31 March 1950)

Astbury's studies of α -keratin, and X-ray studies of crystalline haemoglobin and myoglobin by Perutz and Kendrew, agree in indicating some form of folded polypeptide chain which has a repeat distance of about 5.1 Å, with three amino-acid residues per repeat. In this paper a systematic survey has been made of chain models which conform to established bond lengths and angles, and which are held in a folded form by N—H—O bonds. After excluding the models which depart widely from the observed repeat distance and number of residues per repeat, an attempt is made to reduce the number of possibilities still further by comparing vector diagrams of the models with Patterson projections based on the X-ray data. When this comparison is made for two-dimensional Patterson projections on a plane at right angles to the chain, the evidence favours chains of the general type proposed for α -keratin by Astbury. These chains have a dyad axis with six residues in a repeat distance of 10.2 Å, and are composed of approximately coplanar folds. As a further test, these chains are placed in the myoglobin structure, and a comparison is made between calculated and observed F values for a zone parallel to the chains; the agreement is remarkably close taking into account the omission from the calculations of the unknown effect of the side-chains. On the other hand, a study of the three-dimensional Patterson of haemoglobin shows how cautious one must be in accepting this agreement as significant. Successive portions of the rod of high vector density which has been supposed to represent the chains give widely different projections and show no evidence of a dyad axis.

The evidence is still too slender for definite conclusions to be drawn, but it indicates that a further intensive study of these proteins, and in particular of myoglobin which has promising features of simplicity, may lead to a determination of the chain structure.

1. INTRODUCTION

Proteins are built of long chains of amino-acid residues. Amino-acids unite to form a chain (figure 1) in which the R groups are of some twenty-three different kinds, varying from a hydrogen atom in glycine to moderately complex groups (e.g. the linked five- and six-membered rings of tryptophan). As examples, insulin of molecular weight 12,000 has, according to Chibnall (1945), 106 residues, in four chains linked by six disulphide bridges of cystine (Sanger 1948). Myoglobin, 17,000, has about 146 residues; haemoglobin, 67,000, has about 580 residues.

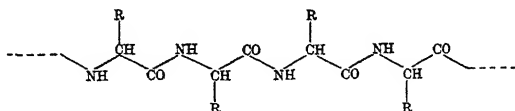


FIGURE 1. Polypeptide chain.

In this paper an attempt is made to glean as much information as possible about the nature of the chains from X-ray studies of crystalline proteins, and to survey the possible types of chain which are consistent with such evidence as is available. The configuration and arrangement of the R groups is not discussed; we are merely

concerned with the configuration of the —CO—CH—NH— chain to which they are attached.

Certain features of all protein structures deserve special mention.

(a) The mean molecular weight of the residue $\begin{array}{c} \text{R} \\ | \\ \text{CH} \\ / \quad \backslash \\ \text{NH} \quad \text{CO} \end{array}$ is much the same in most proteins, ranging between 110 and 120 (e.g. haemoglobin 112.5). Since the molecular weight of the chain element —CO—CH—NH— is 56, one-half the weight of the protein is in the chain and one-half in the side groups, excluding associated water. To put this in another way, the average side-chain contains about four atoms other than hydrogen.

(b) All the amino-acids occurring in proteins (except glycine which is non-enantiomorphous, and a very few amino-acids found in primitive organisms such as bacteria) have the steric configuration about the central carbon atom conventionally termed 'laevo'.

(c) X-ray determinations of the structures of simple amino-acids or dipeptides yield consistent information about the interatomic distances and bond angles in their crystals, and presumably these angles and distances will not be very different in a long polypeptide chain.

2. PREVIOUS SPECULATIONS ABOUT THE CONFIGURATION OF THE
POLYPEPTIDE CHAIN

Astbury and his co-workers in their pioneer investigations have made an exhaustive study of the fibrous proteins such as the keratin of hair and wool. Their most important result, in the present connexion, is their inference that the marked

5.1 Å repeat along the fibre axis which is shown prominently by X-ray photographs of α -keratin and its analogues corresponds to an element of folded chain containing three amino-acid residues. Briefly, Astbury (private communication, 1949) summarizes the evidence as follows:

(a) The β -keratin X-ray diagram with its strong meridional 3.4 Å reflexion represents a system of extended polypeptide chains as in figure 2(b), for which the average length per residue must be about 3.4 Å to correspond to the accepted bond lengths and angles.

(b) The β configuration is approximately twice as long as the folded α configuration, since the reversible extensibility of wool and hair is approximately to twice the normal length. The average length per residue in the α chain is therefore 1.7 Å, giving three residues per repeat of 5.1 Å

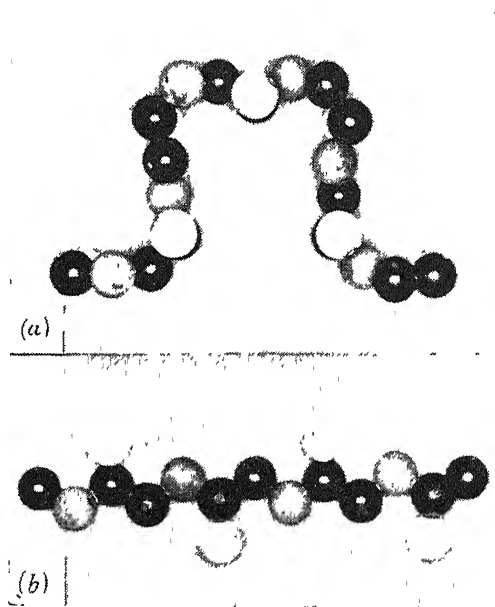


FIGURE 2. Chain configurations proposed by Astbury (1949a) for (a) α -keratin, (b) β -keratin.

The model of the folded chain in figure 2(a) is proposed for α -keratin by Astbury as that which at present seems to fit the facts best, other indications in favour of this structure being obtained from considerations of side chain packing. The complete repeat is at distances of 10.2 Å, each loop occupying 5.1 Å and containing three residues as indicated in the diagram. It will be referred to again below.

Huggins (1943) has made an extensive review of possible types of polypeptide chains. He gives numerous examples of chains in both extended and shrunk forms, and of ways in which these may be linked together, showing how many the possibilities are until further evidence is obtained which restricts them.

Recently a type of chain has been proposed by Ambrose & Hanby (1949; see also Ambrose, Elliott & Temple 1949) based on measurements of the dichroism of

α -keratin, myosin and tropomyosin in the infra-red. This is a folded chain with two residues in a repeat of 5.1 Å, and Ambrose and his collaborators picture these chains as occurring in pairs so that there are four residues in the α -keratin unit.

3. POLYPEPTIDE CHAINS IN CRYSTALLINE HAEMOGLOBIN AND MYOGLOBIN

The X-ray diffraction pictures given by haemoglobin, in particular horse methaemoglobin, have been studied in detail by Perutz and his collaborators (Boyes-Watson, Davidson & Perutz 1947; Perutz 1949), and Kendrew (1950) has published an account of work with myoglobin. The evidence for the existence and features of chains in haemoglobin and myoglobin has been very fully discussed in these papers, but it may be useful to recapitulate here the nature of the problem and the way in which an attack on it has been made. Protein crystals give a wealth of diffracted beams, extending out to angles which correspond to spacings as small as 2 Å, but the interpretation of these photographs as a complete picture of the arrangement of the thousands of atoms in the unit cell is a task greater by several orders of complexity than the most complex crystals yet successfully analyzed. It is always possible, however, to represent the experimental observations as a Patterson or 'vector' map. Broadly speaking, if there are in the actual crystal two atoms a and b with 'weights' m_a and m_b , the Patterson map has a peak or lump of density proportional to the product $m_a m_b$ at a point ' ab ' such that the line joining ' ab ' to the origin of the map is equal and parallel to the line drawn between a and b in the crystal. If the crystal contains n atoms, there are n^2 such points in the Patterson map corresponding to the product $(a+b+c+d+\dots)(a+b+c+d+\dots)$. The n terms $a^2, b^2, c^2, d^2, \dots$ have zero vectors and superimpose at the origin to form a large concentration. The remainder are distributed throughout the cell. (The meaning of Patterson maps is explained in greater detail by Kendrew & Perutz (1949).) It is common practice in crystal analysis to start by forming two- or three-dimensional Patterson summations and then to seek to decipher the significance of their more prominent peaks in terms of the atomic arrangement. However, the Patterson projection of a protein represents some millions of vector peaks. It can only be deciphered if there is some regular underlying arrangement throughout the crystal structure, such as would be the case if the polypeptide chains are regular in form, straight, and parallel to each other in the molecule. It has already been remarked that half the atoms (hydrogen can be neglected on account of its low scattering power) are in the chain element CO—C—N. Further, if some model of the chain is adopted, the first carbon atom of each side-chain is also in a definite position, so that some five-eighths of all the atoms form part of a regularly recurring pattern. The remainder are in the side groups and will have a wide variety of configurations. The vectors drawn between atoms inside any one chain, and those between its atoms and the atoms of neighbouring parallel chains, will form a large proportion of all the vectors in the Patterson and will be of a relatively few constantly repeated types. Owing to this regularity they may be prominent in the Patterson projection in spite of the irregular welter of other vectors on which they are superimposed. This possibility is the greater because the parallel straight chains would be regions of relatively high density. The atoms in any one

chain are bound by homopolar links. The side groups R are in general separated from each other by the much greater distances characteristic of van der Waals forces or polar forces. A simple calculation (Bragg 1949) shows that, in a projection of the crystal structure on a plane perpendicular to the chains, the density of scattering matter immediately round the central axis of a chain may be as much as ten times as great as the average density. To sum up, although it is impossible by trial and error methods to determine the positions of all the atoms in the structure, it may be possible to treat the chains as relatively few units which stand out from the rest of the structure because of their greater density, and to deduce their positions as we would those of atoms in a simpler structure. The first task in protein analysis is therefore to establish the nature of this chain unit.

Figure 3 illustrates the evidence afforded by the haemoglobin and myoglobin analyses. Figure 3(a) is a section of a three-dimensional Patterson projection of haemoglobin, perpendicular to the b axis of the monoclinic crystal and at the height of the origin, which appears half-way along the left-hand side of the figure. It shows a somewhat irregular but clearly marked rod of density through the origin and parallel to the a axis, such as would be produced by the vectors between atoms of polypeptide chains running in this direction. The prominent ring around the origin is at a distance of about 5 Å, and there are further peaks at roughly this spacing along the rod. In figure 3(b) is plotted the total vector weight in a strip along the rod which is 5.2 Å in width, and 3.2 Å in depth in the b direction. It shows a periodicity with a repeat of rather over 5 Å. The positions of the peaks are plotted in figure 3(c); the mean spacing is 5.1 Å.

Sections at other levels in the haemoglobin cell show more chains parallel to the a axis, in positions which indicate that each chain is surrounded by six others at distances of 10 to 11 Å. Assuming this type of packing, the number of residues in each repeat of 5.1 Å can be calculated from the density of the protein (1.3) and the mean residue weight (112.5), and is found to be 3.3 residues per unit of repeat. This may be compared with Astbury's conclusion that in α -keratin there are three residues in a repeat of 5.1 Å. No great reliance can be placed on the figure deduced from haemoglobin, since the estimates of packing distance and repeat distance are very approximate, and in particular the possibility of four residues per 5.1 Å repeat cannot be regarded as ruled out.

Figure 3(d) is a two-dimensional Patterson projection of myoglobin on the b plane of the monoclinic crystal. It is not to be expected that chains should stand out so clearly in a two-dimensional projection as in one of three dimensions; nevertheless, there is a clear indication of rods running in the $[20\bar{1}]$ direction. Peaks appear along the chains at intervals of about 10 Å. In a Patterson which includes a larger number of higher-order terms, intermediate peaks appear at intervals of about 5 Å. The data from myoglobin are discussed more fully in a later section, where a further Patterson projection is described in which the rods are seen in an end-on direction. The evidence from the b projections suggests the existence of parallel chains with a repeat of 10 Å and a subsidiary repeat of 5 Å.

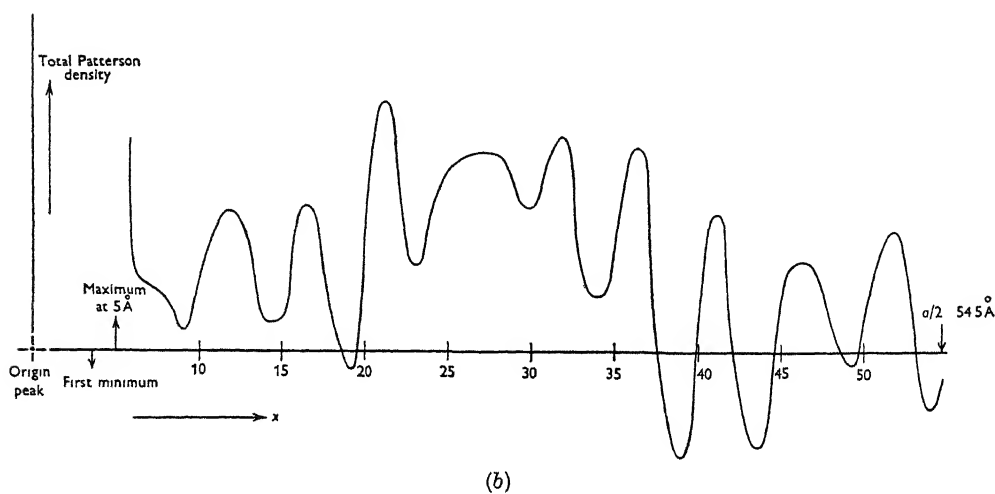
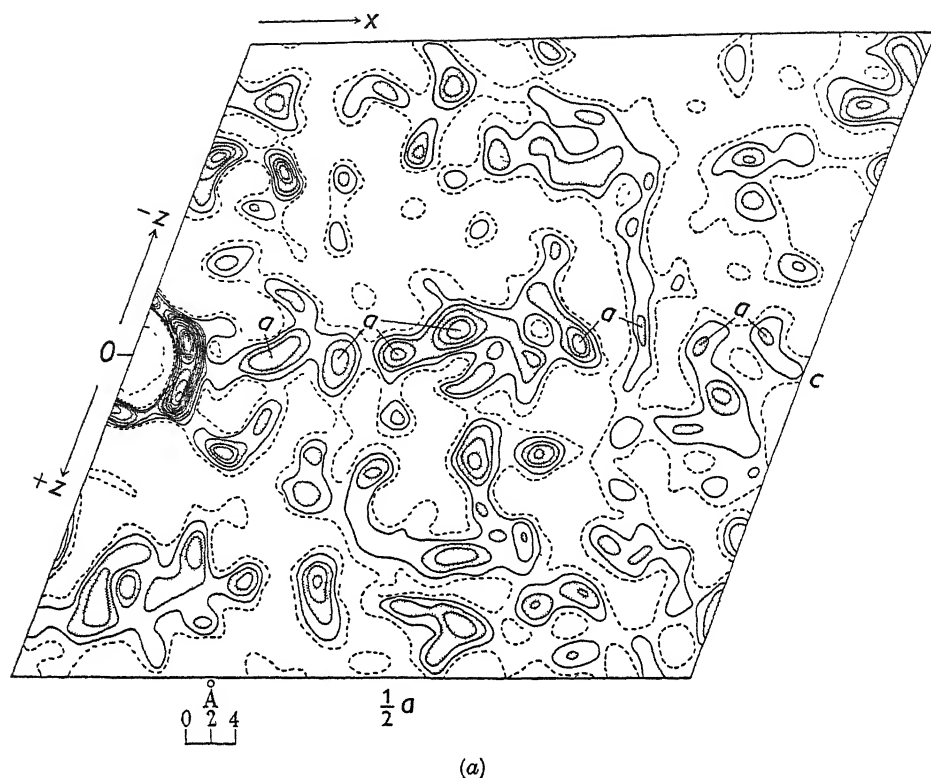
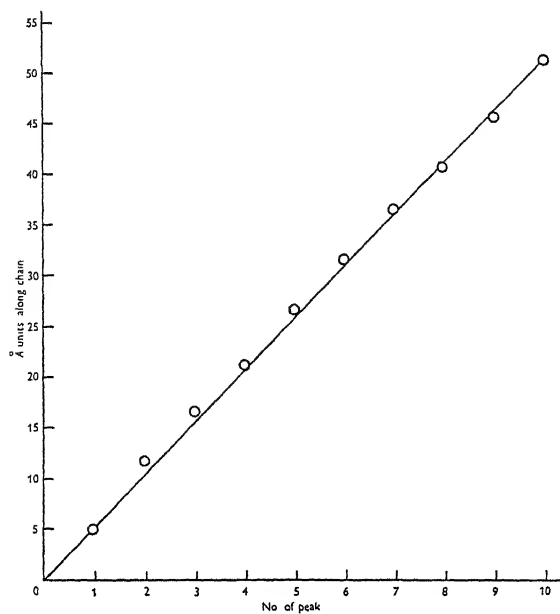
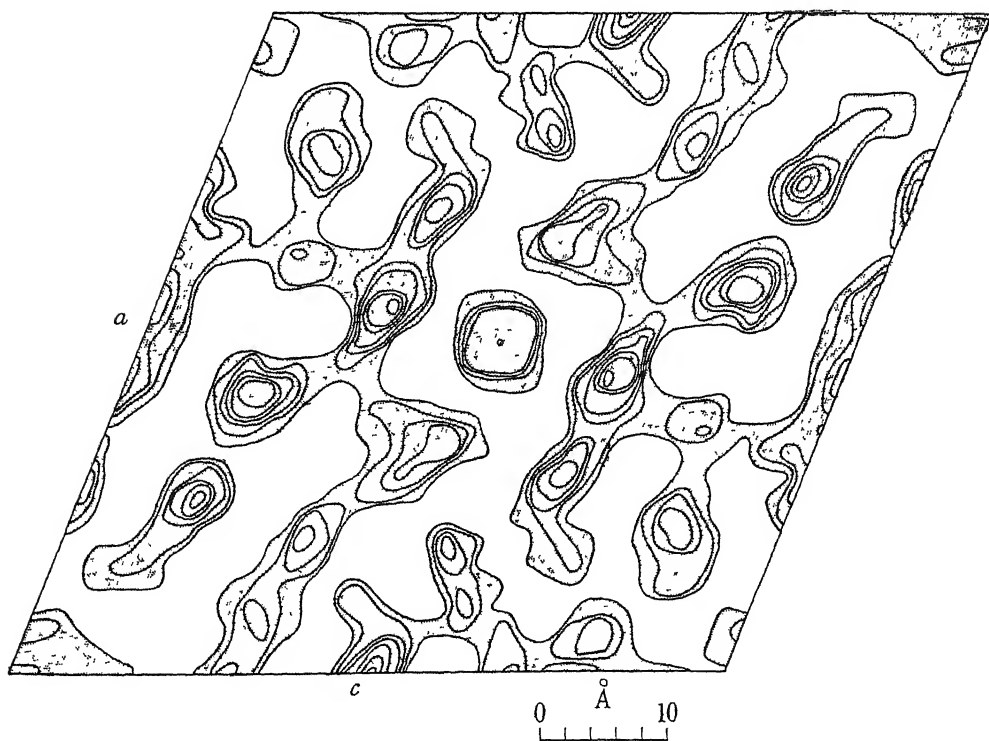


FIGURE 3. (a) Patterson section of horse methaemoglobin taken through the origin and perpendicular to b , showing vector rod parallel to a . (b) Integrated vector density in a strip of the Patterson section of (a) including the rod, plotted as a function of distance from the origin. (c) Positions of vector peaks in (b) plotted against ordinal number of peak. (d) Patterson projection of horse met-myoglobin perpendicular to b , showing vector rods parallel to $[201]$ (from Kendrew 1950).



(c)



(d)

4. CLASSIFICATION OF CHAIN STRUCTURES

In this and the following sections we attempt to survey systematically all those types of folded polypeptide chain configurations which satisfy certain conditions, established by experiment or plausible on general grounds.

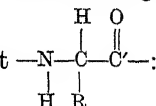
It cannot be assumed as certain that the polypeptide chain has the same configuration in all crystalline proteins, or that a similar configuration occurs in fibrous proteins such as α -keratin. It is, however, not unreasonable to expect that haemoglobin and myoglobin contain chains of the same type, because these proteins appear to be closely related in several ways; and furthermore, the repeat distance, the inter-chain distance, and the number of residues per repeat, are similar in these two proteins to the corresponding features of α -keratin. It will therefore be assumed as a working hypothesis that the chain configurations in large classes of proteins resemble one another closely, while bearing in mind that this hypothesis is based on slender evidence and may have to be abandoned when further experimental data are available.

In our survey of chain configurations we have adopted the following conditions:

(a) Interatomic distances and bond angles

Some relevant data have been obtained by X-ray analysis of the structures of amino-acids and small peptides. Glycine has been analyzed by Albrecht & Corey (1939), DL-alanine by Levy & Corey (1941) and β -glycylglycine by Hughes & Moore (1949). An addition compound of cysteyl-glycine has recently been analyzed in the Cavendish Laboratory by Dyer, and progress has been made with the analysis of tripeptides and tetrapeptides.

In an earlier paper Huggins (1943) has given the following summary of the distances and angles to be expected in the chain unit



	(Å)		(°)
N—C	1.41	$\begin{array}{c} \diagup \\ \text{NCC}' \\ \diagdown \end{array}$	112
C—C'	1.52	$\begin{array}{c} \diagup \\ \text{CC}'\text{N} \\ \diagdown \end{array}$	118
C'—O	1.25	$\begin{array}{c} \diagup \\ \text{C}'\text{NC} \\ \diagdown \end{array}$	118
C'—N	1.33		

The experimental results, including the most recent, agree generally with these figures, though with some individual variations (for summary see Corey 1948). For example, in cysteyl-glycine Dyer has found $\text{C}'\text{NC}$ considerably greater than the figure given. All the structures contain hydrogen bonds $\text{N—H}\cdots\text{O}$ which vary in length between 2.6 and 2.85 Å.

For the construction of models we have adopted the following values, which agree with those given by Huggins or with generally recognized standards:

- (i) Covalent distances: C—C, 1.52 Å; C—N, 1.36 Å; C—O, 1.24 Å.
 (ii) Hydrogen bond distances: N—O (in N—H...O) 2.85 Å.
 (iii) Bond angles: C. tetrahedral distribution.
 N: tetrahedral (interbond angle 109° 28') or planar (120°).*
 N—H.. O=C<: we have placed no restriction on $\widehat{\text{NHO}}$ or on $\widehat{\text{HOC}}$, though we have generally attempted to make NHOC as nearly collinear as possible.

(b) Optical configuration of amino-acids

We have assumed that all the amino-acids (except glycine which is optically inactive) belong to the *laevo* series.

(c) Symmetry of the chain

Huggins (1943) has made the following two general points about the symmetry of a stable chain configuration:

(i) 'Polypeptide chains extending through the crystalline regions must each have a screw axis of symmetry, or else two or more chains must be grouped around screw axes or other symmetry elements. The unbalanced forces on opposite sides of a chain which has no screw axis, e.g. any of the earlier chain structures advocated for α -keratin by Astbury or the one that he has most recently proposed for collagen, would tend to bend it continuously in the same direction.'

(ii) 'In general, a structural pattern for a protein in which like groups are all surrounded in a like manner, except for differences between the R groups, is more probable than one in which this is not the case.'

We have accepted Huggins's first criterion throughout; indeed, we have used it as a basis for classifying types of configuration. In other words, all the structures examined possess a screw axis of symmetry (not necessarily restricted to crystallographic types of screw symmetry, e.g. a fivefold axis would be permissible).

Huggins's second criterion, that each element of the chain should be in a similar relation to neighbouring elements, we have not regarded as essential, and some of the chains described below do not obey this rule.

(d) The role of hydrogen bonding

We have made the plausible (but still unproved) assumption that the chain is held in a folded condition by hydrogen bonds between >NH and >CO groups of nearby amino-acid residues. In other words, the folded chain is thrown into a series of rings which must be ruptured at the hydrogen bond before unfolding can take place. Formally speaking, these rings may be of two types, illustrated diagrammatically in figure 4(a).

* No complete structure analysis of any compound containing nitrogen bound analogously to nitrogen in proteins of the α -keratin type is yet available. We have therefore thought it better to leave this question open for the present, and to examine, in each type of structure, the effect of the two configurations given.

We have used criteria (c) and (d) as a basis for classifying possible types of chain configuration:

(i) *By symmetry*. The chains may possess a twofold, threefold, fourfold or higher screw axis of symmetry.

(ii) *By ring size*. The ring, being formed by elements of the chain and the hydrogen atom between O and N, must be of one of the following types:

Type A. Number of atoms in ring, $R = 3n + 4 = 7, 10, 13, \dots$

Type B. Number of atoms in ring, $R = 3n + 5 = 8, 11, 14, \dots$

These rings are shown schematically in figure 4(b).

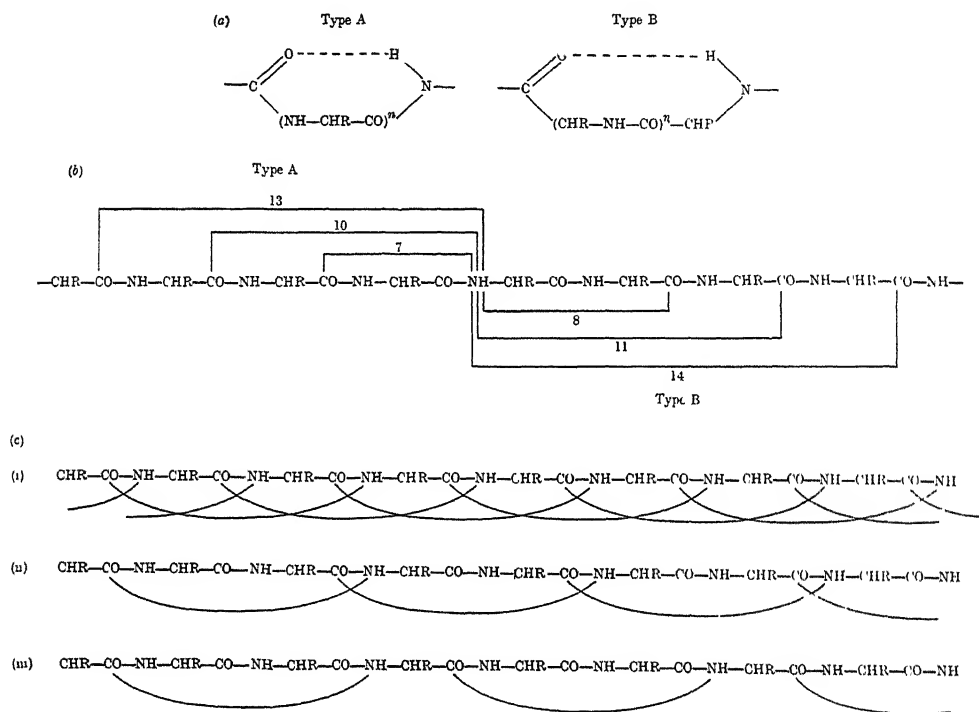


FIGURE 4. (a) Possible types of hydrogen-bonded ring formed from folded polypeptide chain.

(b) Scheme showing the numbers of atoms in the various possible ring structures.

(c) Families of structures of type S_{10} : (i) $1/q = 1$, (ii) $1/q = \frac{1}{2}$, (iii) $1/q = \frac{1}{3}$.

Each type of structure may be designated by a symbol S_R , where S denotes the screw symmetry and R the number of members in the hydrogen-bonded rings; e.g. 2_7 is a structure having a dyad axis of symmetry with seven-membered rings. Formally speaking each of these structural types may be further subdivided into families according as all, or only a fraction $1/q$, of the NH (or CO) groups form hydrogen bonds. For example, the structure type S_{10} might form families of the kind shown in figure 4(c); in (i) every possible hydrogen bond is made, in (ii) $\frac{1}{2}$ and in (iii) $\frac{1}{3}$ of the possible bonds are made. Similar families can be described for each of the permitted values of R . These may be designated by the symbols $S_R(1/q)$, e.g. $2_7 \cdot \frac{1}{2}$ would be a structure with a seven-membered ring, a dyad screw axis, and one-half of the possible hydrogen bonds made.

It will be evident that the number of configurations formally possible is rather large, and even though many of the formal schemes are found on examination not to be sterically possible it is still true that the number of configurations which should be examined in a comprehensive survey is considerable. We regard those structures in which all NH and CO groups are hydrogen bonded ($q = 1$) as inherently the more probable, because their free energy is presumably lower. We have therefore examined structures of this type in considerable detail and have made efforts to build all possible models provided they conformed to the conditions outlined above.

We have also tried to build all the possible structures having only a fraction of the NH and CO groups hydrogen bonded, determining at least the repeat distance of the molecular pattern and the number of amino-acid residues which that pattern contains. This survey was less thorough, so that some possible structures may have been overlooked, and only a few of the structures were examined in detail. Among the latter was $2_{13} \cdot \frac{1}{3}$ because of the evidence which Astbury produced in its favour, and $2_{14} \cdot \frac{1}{3}$ because of its close similarity with $2_{13} \cdot \frac{1}{3}$.

5. STRUCTURES EXAMINED IN DETAIL

Within the limits discussed in the last two paragraphs we have attempted to make models of as many types of structure as possible, considering in turn each type of chain symmetry and each ring size. In some instances no structure could be devised for steric reasons. In general, very small or very large rings may be excluded; small rings on account of the strains involved in closing them, large ones because they involve either excessively long hydrogen bonds or else chains of very large cross-sectional area containing many residues per repeating unit.

These structures examined in detail are listed in table 1, classified in accordance with their symmetry and ring size. It is to be noted that a given chain configuration can in general give rise to two types of structure depending on the way the side groups are attached, since the chain itself is enantiomorphous; broadly speaking, these may be distinguished as *laevo* groups in a left-handed chain, or *laevo* groups in a right-handed chain—structures which are *not* mirror images of each other and which are designated by suffixes 'a' and 'b'. In some cases steric considerations rule out one of these configurations.

We now proceed to discuss briefly each of the structures listed in table 1, leaving over for more detailed treatment later those which we regard as displaying the most promising agreement with the experimental data. Those not further considered have generally been excluded because they do not satisfy one or both of the following criteria:

(a) *Repeat distance.* The structure must account for an apparent 'repeat' distance of just over 5 Å. This 'repeat' may be the true repeat distance between identical points, or a fractional translation produced by the action of the screw axis.

(b) *Number of residues per repeat.* We have preferred structures in which the repeat (or pseudo-repeat) contains three amino-acid residues, for the reasons given above (p. 325). We do not feel, however, that we can entirely exclude the possibility that the repeat contains four residues, and two such structures are considered in

TABLE 1

screw axis of symmetry	no. of atoms in ring	repeat distances (Å)	no. of residues per repeat	illustra- tions (figure no.)	comments
twofold	7a	5-5.6	2	5	structure proposed by Huggins (1943)
	7b	5-5.6	2	6	structure proposed by Zahn (1947) and Ambrose <i>et al.</i> (1949); readily folds in pairs, see §7, 8
	8	4.6-4.8	2	7	structure proposed by Huggins (1943); only one configuration possible; repeat distance too short
	13	10.2	6	8	structure proposed by Astbury & Bell (1941); see §7, 8
	14	10.2	6	9	see §7, 8
threefold	7	7.5	3	—	repeat distance too short
	8	5.4	3	10	hydrogen bonds mutually perpendicular; see §7
	10	5.2	3	11	structure proposed by Taylor (1941) and Huggins (1943); hydrogen bonds oriented nearly parallel to the chain direction; see §8
	11	—	—	—	} no possible structures
	13	—	—	—	
fourfold	14	—	—	—	
	7	—	—	—	} no possible structures
	8	—	—	—	
	10	—	—	—	
	11	5.4	4	—	rings somewhat strained; similar to 4 ₁₃ ;
	13	5.6	4	12	a possible structure
	14 or greater	—	—	—	no possible structures
fivefold and higher symmetries	—	—	—	—	all such structures contain more than four amino-acid residues per repeat unit

detail (2_{7a} and 4₁₃). It appears highly improbable that the number of residues in the 5 Å repeat is greater than four, and this criterion excludes symmetries which are more than fourfold; a fivefold chain repeating at 5 Å, for example, would necessarily contain at least five amino-acid residues per repeat.

The following diagrams illustrating the types of chain were drawn from optical projections of models, and are only intended to illustrate the structure of the chain. The co-ordinates of the atoms given in the accompanying tables are accurate to about 0.1 Å.

(i) *Twofold structures*

Seven-atom ring (2_{7a}, 2_{7b})

One form of this ring (2_{7a}) was proposed by Huggins (1943), whose drawings of it are reproduced in figure 5. The chain itself forms a more or less flat ribbon while R groups project alternately up and down from the plane of the ribbon. The repeat distance is 5 to 5.6 Å and contains two residues.*

* The precise repeat distances in 2₇ and 2₈ cannot be predicted since the N—H—O—C bonds deviate widely from a straight line, and precise experimental evidence is not yet available for the N—O distance in such a case.

The other form (2_7b), differing only in that the R groups and H atoms on the α -carbon atoms are interchanged, was suggested by Zahn (1947) and described in more detail by Ambrose & Hanby (1949) and Ambrose *et al.* (1949). As shown in figure 6 the R groups now lie approximately in the plane of the ribbon, instead of normal to it. This is an attractive feature of the structure, since it enables the chain

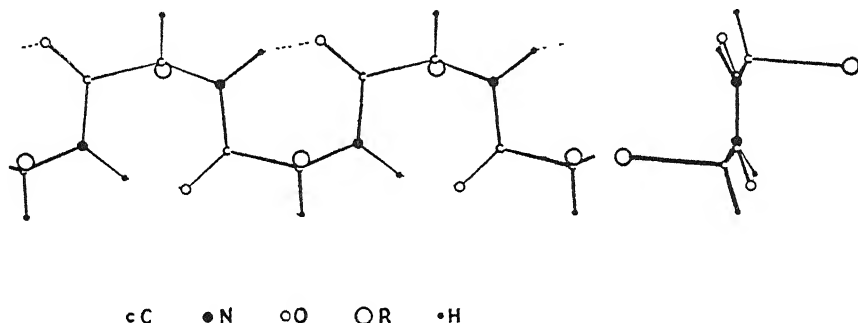


FIGURE 5. Projections of 2_7a chain, after Huggins (1943). Repeat distance about 5 Å

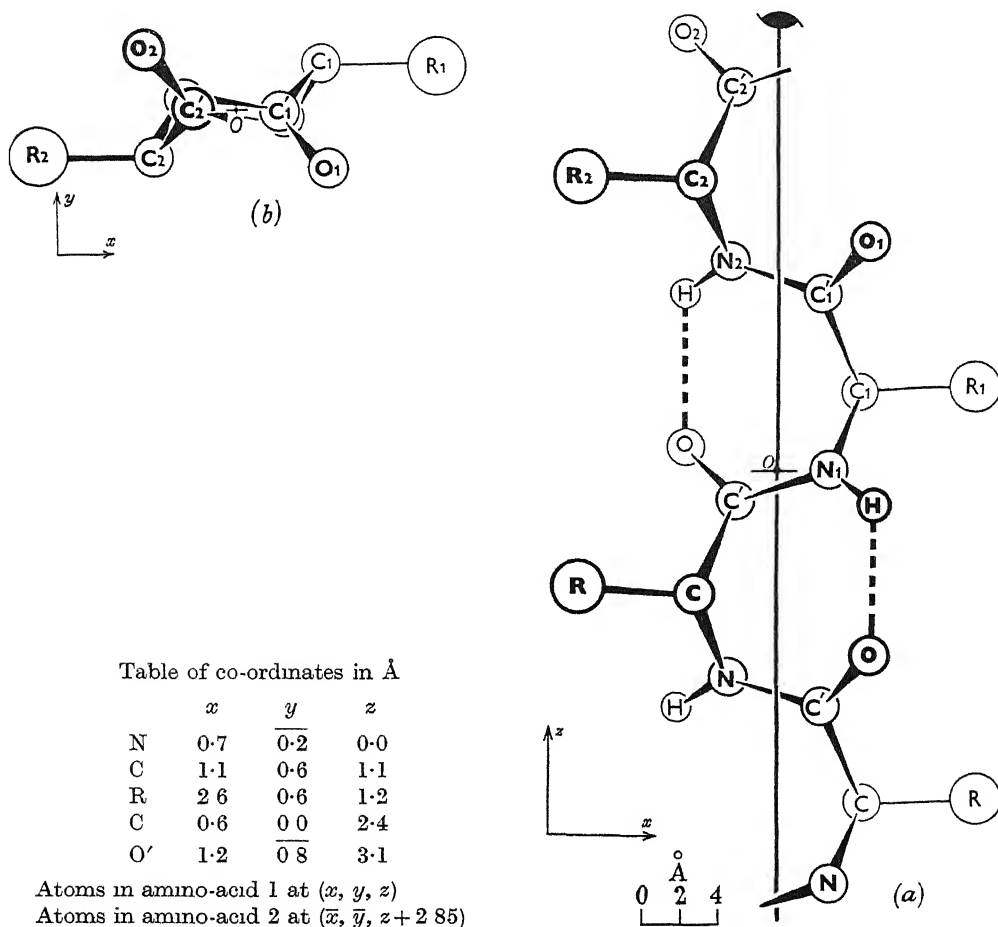


FIGURE 6. Projections of 2_7b chain.

to bend back on itself about an axis in the ribbon plane, without mutual interference of R groups or rupture of hydrogen bonds; the two limbs then lie about 4 Å apart.

Astbury (1949*b*) has criticized 2_7b on several grounds, one being that a chain having a true repeat of 5 Å and a screw dyad axis would give a weak 010 (5 Å) and a strong 020 (2.5 Å) reflexion, whereas in fact in α -keratin it is the 5 Å meridional reflexion which is overwhelmingly strong. This objection would apply to both forms of 2_7 , and to overcome it Huggins (1943) introduced 'the additional assumption that alternate R groups are much more potent X-ray scatterers than the intermediate ones'. There is, however, no other evidence that such an arrangement exists.

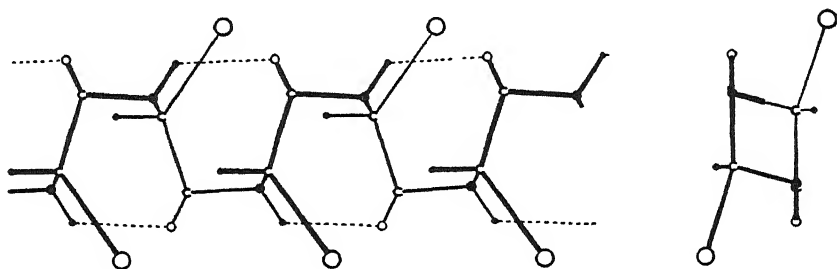


FIGURE 7. Projections of 2_8 chain after Huggins (1943). Repeat distance about 5 Å. Symbols as for figure 5.

2_7b has also been criticized by Darmon & Sutherland (1949) on grounds connected with the interpretation of infra-red absorption spectra given by Ambrose *et al.* (1949).

Another serious objection to both forms of 2_7 is that it contains only two amino-acid residues per repeat of 5.5 Å. Alternatively, 2_7b can be arranged so that two close-spaced chains (see above) run through a single subcell of $5 \times 10 \times 10$ Å, giving *four* residues per subcell. We have given reasons for preferring models containing *three* residues per subcell, but these reasons are not conclusive ones, and so we have given more detailed consideration (see below) to 2_7b , in particular in relation to haemoglobin and myoglobin.

Both structures can be built using either a planar or a tetrahedral configuration of bonds about the nitrogen atom. We have preferred a planar configuration, since this gives a more nearly linear arrangement of the hydrogen bond N—H—O, and because the ring is flatter than with the tetrahedral configuration, which prevents the packing of neighbouring chains to as close a distance as the myoglobin data would appear to demand (see p. 350).

Eight-atom ring (2_8)

This model was first suggested by Huggins (1943), whose projections of it are reproduced in figure 7. In the configuration of its R groups it resembles 2_7b rather than 2_7a ; no second form, corresponding to 2_7a , is possible for steric reasons; in such a structure R would lie far too close to C'.

The 2_8 structure is not immediately attractive since, like 2_7 , it contains only two residues in the repeat distance and has other disadvantages not shared by 2_7 :

(a) The repeat distance is distinctly less than 5 Å; 4.6 with planar, and 4.7 to 4.8 Å with tetrahedral nitrogen atoms.

(b) The angle HOC in the hydrogen bonds is only about 100° , which seems unlikely on general grounds, since it brings the N and O atoms very close to one another.

(c) The whole chain is much more rigid than 2_7b , leaving no possibility of folding about an axis in the plane of the ribbon. In any case the side chains would interfere with such a fold.

For all these reasons we regard 2_8 as an unlikely structure, and do not consider it further.

Thirteen-atom ring ($2_{13} \cdot \frac{1}{3}$)

This is the well-known structure proposed for α -keratin by Astbury & Bell (1941). We have constructed models in conformity with the diagrams published by these authors; our projections are illustrated in figure 8. The repeat distance is just more than 10 \AA , and contains six amino-acid residues; side chains project alternately up and down. The structure satisfies all the experimental data and is discussed in more detail below.

We found it impossible to construct a $2_{13} \cdot \frac{1}{3} \cdot b$ configuration. On the other hand, we obtained a quite distinct version of $2_{13} \cdot \frac{1}{3}$ in which the side-chains project alternately up and down as in the Astbury version, and the repeat distance is 9 \AA , containing six amino-acid residues. The ring is somewhat strained and the structure seems in general an unlikely one, we mention it merely as an illustration of the fact, which we have observed here and in $2_{14} \cdot \frac{1}{3}$, that these large ring structures have considerable flexibility and their detailed configurations can be altered within wide limits. At the present stage there is clearly nothing to be gained by studying all such minor variations in detail.

Fourteen-atom ring ($2_{14} \cdot \frac{1}{3}$)

This structure, which we believe has not been previously described, is illustrated in figure 9. It possesses all the attractive features of $2_{13} \cdot \frac{1}{3}$: the ring is not strained and is almost planar: R groups project alternately up and down from the ring plane (though they are not close-packed, their distances apart varying from 4 to 6 \AA); the repeat distance is 10.2 \AA and contains six residues. Accordingly, we discuss the structure in detail below.

Both $2_{14} \cdot \frac{1}{3} \cdot a$ and $2_{14} \cdot \frac{1}{3} \cdot b$ are possible, and each is susceptible of minor variations, but the ring is so flexible that we have only thought it worth while to discuss one typical, and on general grounds highly probable, structure.

(ii) *Threefold structures*

In all these structures (and those of higher symmetry) the R group and H atom on the α -carbon atoms can easily be interchanged, with little effect on the general properties of the structure. We have in each case considered the version in which the side-chains would protrude most nearly normal to the axis of symmetry.

Seven-atom ring (3_7)

This structure would appear to be possible sterically, and we constructed a model without difficulty. We have not considered it further, since its repeat distance is 7.5 \AA (containing three amino-acid residues).

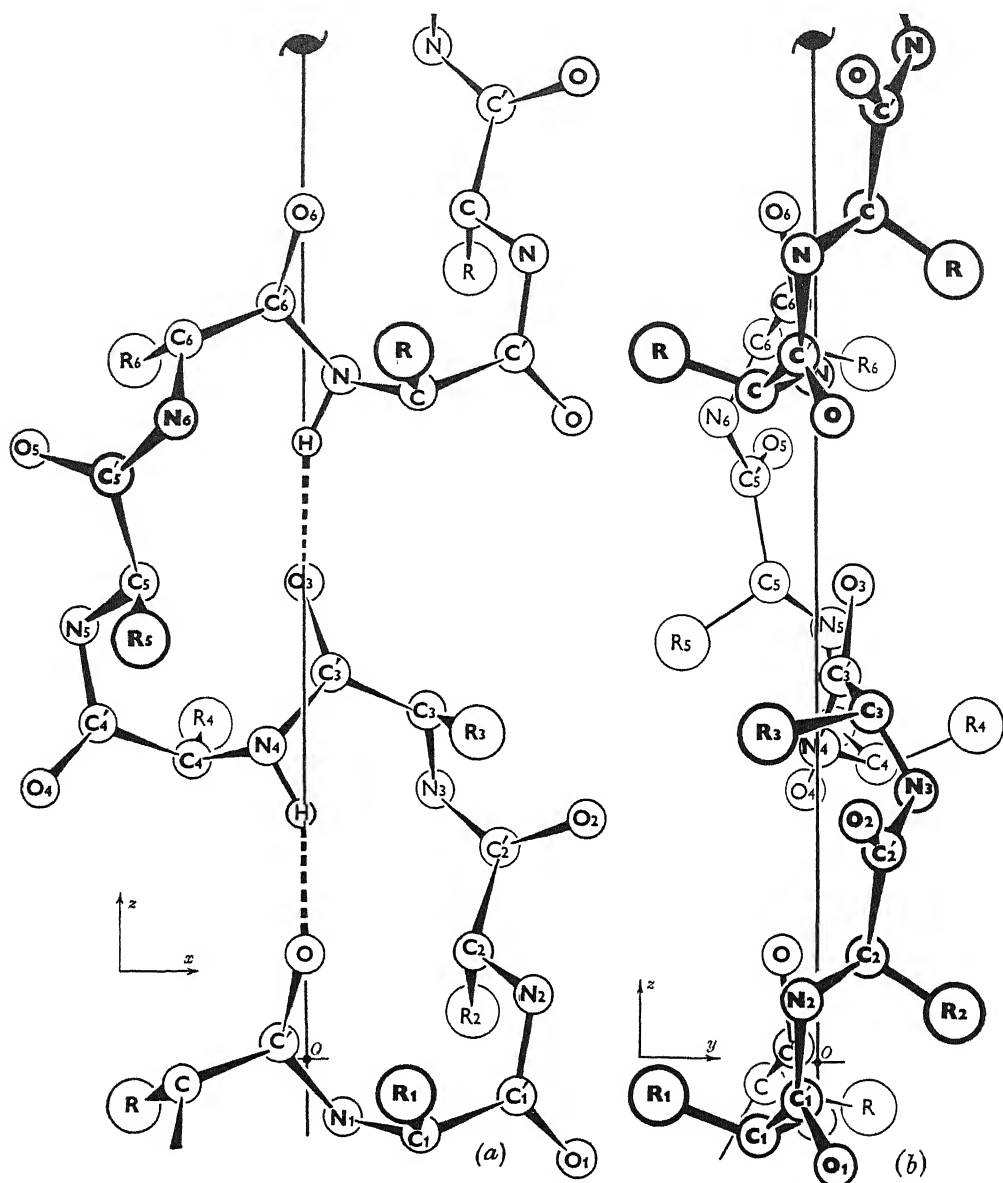


Table of co-ordinates in Å

	x	y	z
N_1	0.5	0.0	0.8
C_1	1.6	0.8	1.0
R_1	1.4	2.2	0.5
C'_1	2.9	0.2	0.5
O_1	3.7	0.2	1.4
N_2	3.1	0.2	0.9
C_2	2.3	0.7	1.5
R_2	2.3	1.9	0.7
C'_2	2.7	0.9	2.9
O_2	3.8	0.6	3.3
N_3	1.8	1.4	3.7
C_3	1.7	0.7	4.8
R_3	2.4	0.7	4.5
C'_3	0.4	0.3	5.3
O_3	0.0	0.5	6.5

Atoms in amino-acid residues 4, 5, 6 at $(\bar{x}, \bar{y}, z \pm 5.1)$

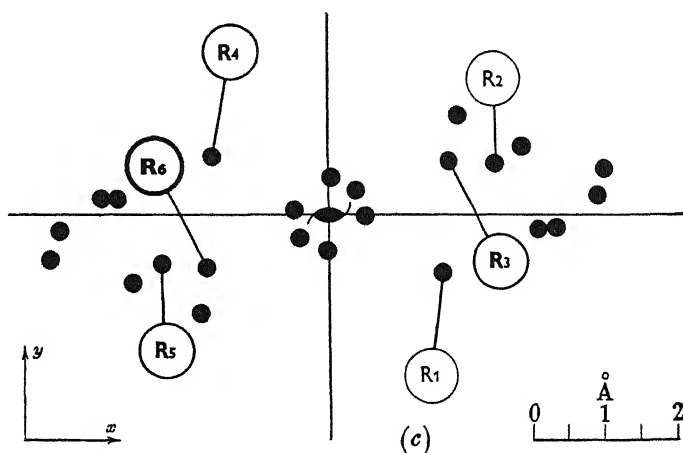


FIGURE 8. Projections of $2_{13}.1$ (Astbury) chain.

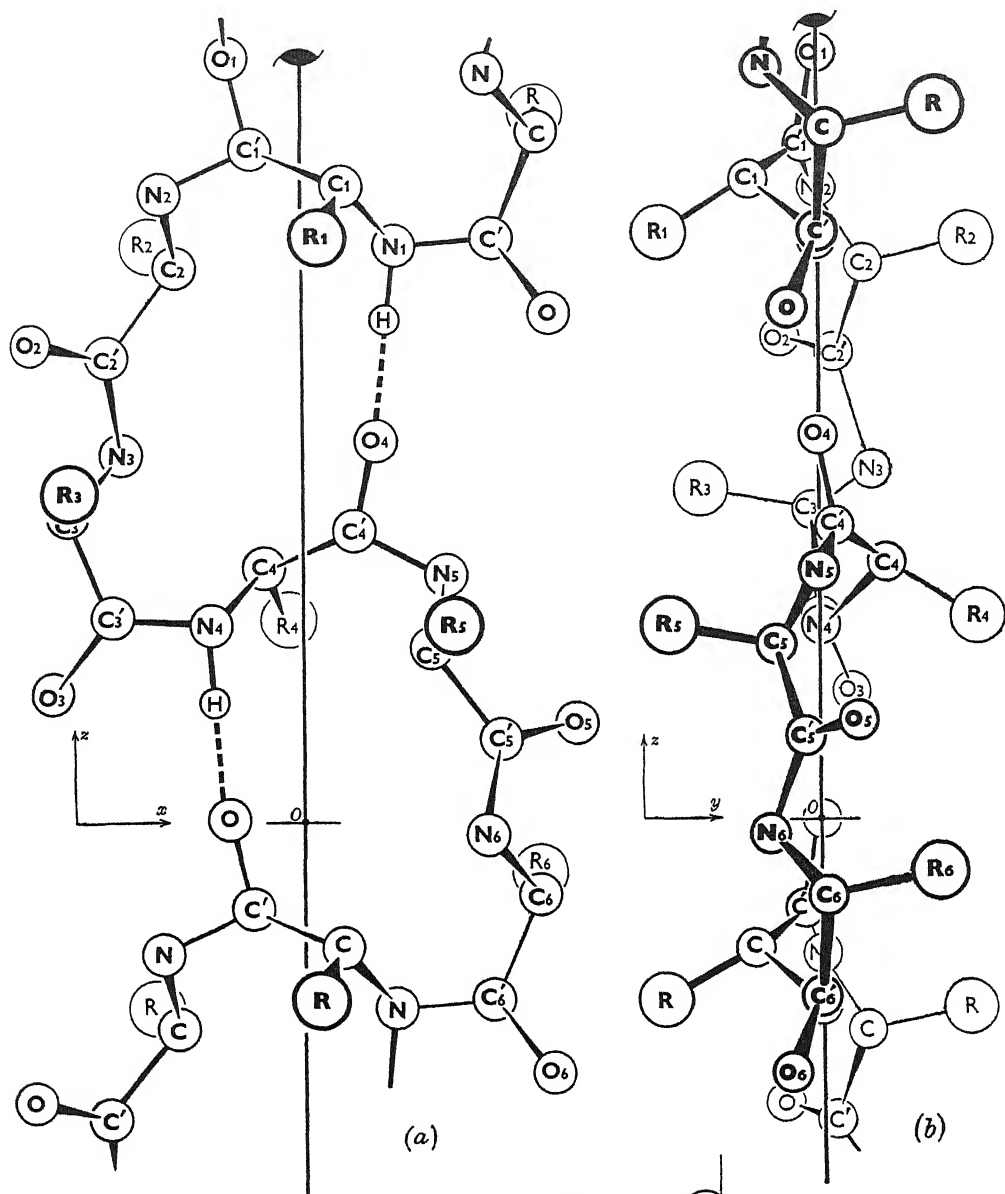


Table of co-ordinates in Å

	1			2		
	<i>x</i>	<i>y</i>	<i>z</i>	<i>x</i>	<i>y</i>	<i>z</i>
N	1.2	0.0	7.7	1.9	0.0	8.4
C	0.5	0.9	8.5	1.7	0.6	7.4
R	0.2	2.1	7.8	2.0	2.0	7.7
C'	0.7	0.2	9.0	2.6	0.2	6.2
O	1.0	0.0	10.2	3.6	0.5	6.4

	3		
	<i>x</i>	<i>y</i>	<i>z</i>
N	2.4	0.7	4.9
C	3.1	0.1	4.1
R	3.1	1.6	4.4
C'	2.5	0.0	2.7
O	3.3	0.4	1.7

Atoms in amino-acid residues 4, 5, 6
at $(\bar{x}, \bar{y}, z \pm 5.1)$

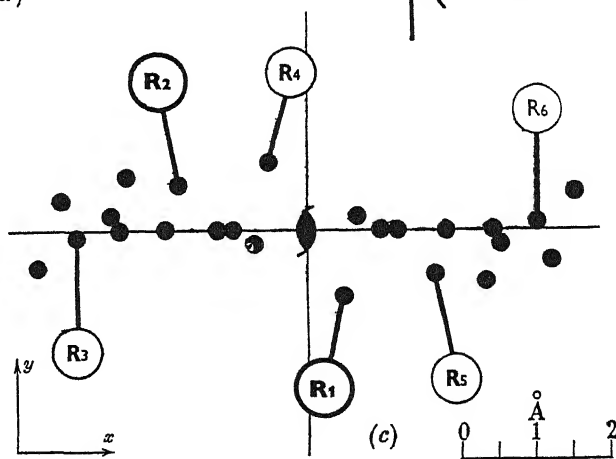


FIGURE 9. Projections of $2_{14} \cdot \frac{1}{3}$ chain.

We have not encountered any earlier description of this structure. Our model of it is illustrated in figure 10. The repeat distance is about 5.4 \AA , and of course comprises three amino-acid residues. It is a feature of this structure that the three

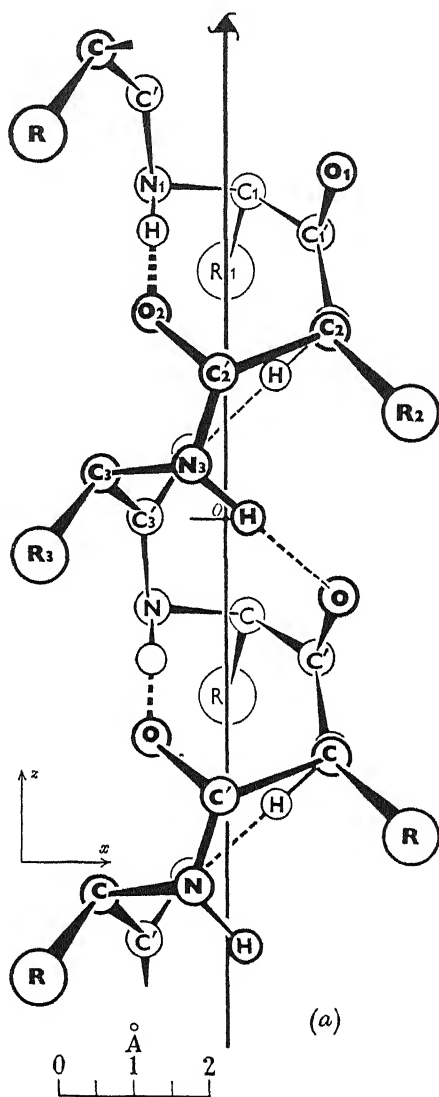


FIGURE 10. Projections of 3_8 chain.

N—H...O bonds in each repeating unit are mutually perpendicular and therefore could not possibly be responsible for any infra-red pleochroic effects which might be observed in the crystal or fibre. However, we regard the whole interpretation of these effects as still equivocal, and hence, though we shall not discuss this model further (owing to its close similarity to 3_{10}), we cannot definitely reject it at this stage.

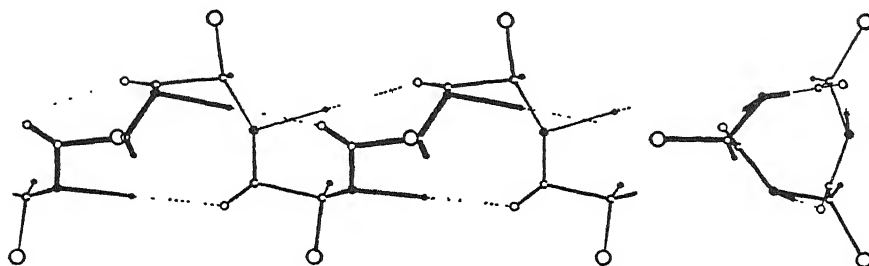


FIGURE 11. Projections of 3_{10} chain, after Huggins (1943). Repeat distance about 5.1 Å. Symbols as for figure 5.

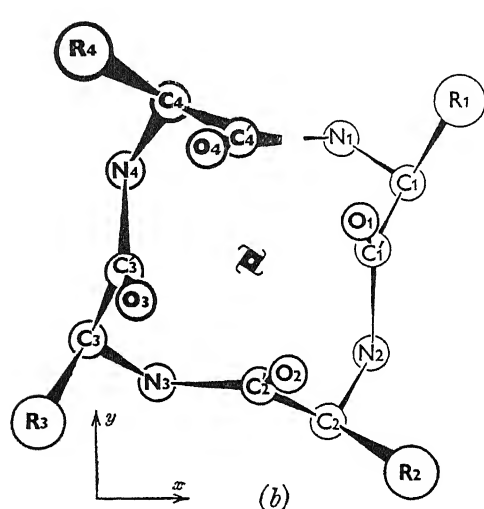


Table of co-ordinates in Å

	x	y	z
N	1.2	1.6	1.0
C	2.1	1.0	1.8
R	2.8	2.1	2.4
C'	1.6	0.1	2.8
O	1.5	0.5	3.9

Atoms in amino-acid 1 at (x, y, z)

Atoms in amino-acid 2 at $(y, \bar{x}, z + 1.4)$

Atoms in amino-acid 3 at $(\bar{x}, \bar{y}, z + 2.8)$

Atoms in amino-acid 4 at $(\bar{y}, x, z + 4.2)$

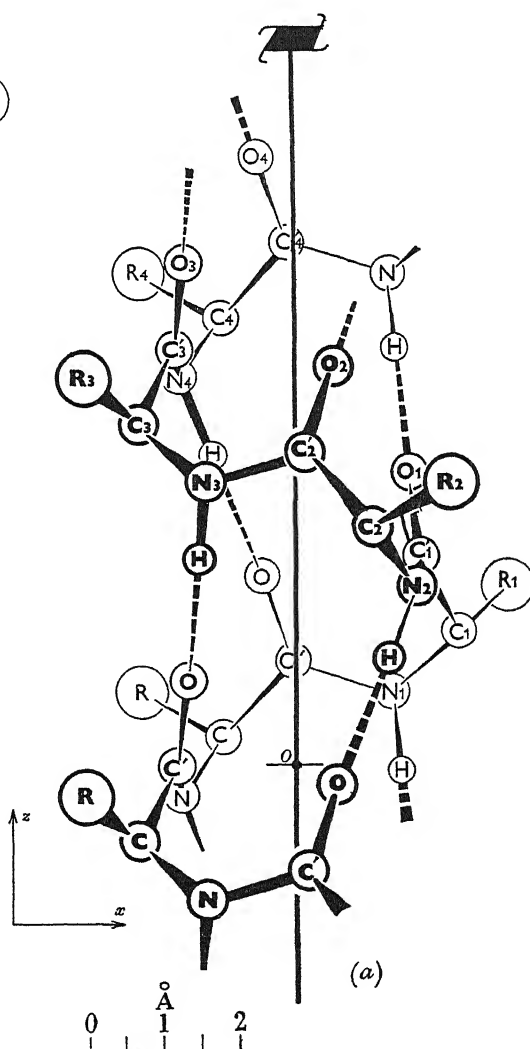


FIGURE 12. Projections of 4_{13} chain.

Ten-atom ring (3_{10})

This structure was first proposed by Taylor (1941), it was later discussed in detail by Huggins (1943), whose projections of it are reproduced in figure 11. Like 3_8 its repeat distance is rather over 5 \AA , and contains three residues; it has the further attractive feature that the hydrogen bonds are oriented nearly parallel to the triad axis and could therefore contribute to the infra-red pleochroism of the molecule as a whole, granted a suitable orientation of the chains.

We discuss this structure in detail below.

Other threefold structures

We have found it impossible to build threefold structures containing rings of 11, 13 or more members; structures with these large rings invariably assume a symmetry higher than threefold.

(iii) *Fourfold structures*

We were able to build no fourfold structures having rings of fewer than eleven or more than thirteen members

Eleven-atom ring (4_{11})

A structure of this type can be built, though neighbouring pairs of oxygen atoms fall abnormally close to one another (2.2 \AA instead of the normal minimum of 2.8 \AA). Because of this feature, and because the repeat unit (of 5.4 \AA) contains *four* amino-acid residues, the structure is regarded as improbable.

Thirteen-atom ring (4_{13})

This version seems to be more plausible than 4_{11} since its rings are less strained; projections are given in figure 12. The repeat distance is again rather over 5 \AA and, of course, contains four amino-acid residues. This structure will be further referred to in §7.

6. STRUCTURES NOT EXAMINED IN DETAIL

(a) *Structures of higher symmetry*

We have not examined in detail structures with fivefold, sixfold or higher symmetry since they would inevitably contain more than four residues per repeat unit. There is, nevertheless, no difficulty in building models of such structures. For example, there is a hexad structure 6_{20} , with a repeat distance of about 5 \AA and containing six residues per repeat; it is an open helix whose internal diameter is between 7 and 8 \AA .

(b) *Structures in which only part of the CO and NH groups are hydrogen bonded*

We mentioned in §4 above that structures of this type were considered comparatively unlikely, because they would be expected to have a higher free energy than those in which all CO and NH groups are bonded. Nevertheless, a general survey has been made. The list of possible structures is not as formidable as might appear at first sight, because the repeat distances along the fibre axis become too long when the fraction of bonded NH or CO groups falls below $\frac{1}{3}$, and often already when it falls

below $\frac{1}{2}$. Some structures were found to be sterically impossible, and others again, though capable of shortening to give the desired repeat of $n \times 5 \text{ \AA}$, contained no bond that would keep them in this shortened configuration.

The structures are described in table 2, which includes only three models having the correct repeat distance. Though these three, as well as some of the other structures listed in table 2, may have any type of screw symmetry owing to the rotational freedom of the individual rings, we decided for simplicity's sake to describe them all in terms of twofold screw axes. We append some comments on the three structures having approximately correct repeat distances.

TABLE 2

screw axis of symmetry	no. of atoms in ring	fraction of NH (and CO) groups which are hydrogen bonded	true repeat	no. of residues per repeat	illustrations (figure no.)	comments
twofold	7	$\frac{1}{2}$	11.6	4	13	two side-chains above and two below plane of main chain; free rotation between rings
	7	$\frac{1}{3}, \frac{1}{4}$, etc.	≥ 10	6, 8, etc.	—	'repeat' too long
	8	$\frac{1}{2}$	9.8	4	14	free rotation between rings
	8	$\frac{1}{3}, \frac{1}{4}$, etc.	≥ 10	6, 8, etc.	—	'repeat' too long
	10	$\frac{1}{2}$	8.2	4	—	'repeat' too short
	10	$\frac{1}{3}$	12	6	—	'repeat' too long
	11	$\frac{1}{2}$	8.0	4	—	'repeat' too short
	11	$\frac{1}{3}$	9.6	6	15	possible structure; properties similar to 2 ₁₃ and 2 ₁₄ , though rings are not so flat; side chains point alternately up and down from the plane of the rings; free rotation between rings
	11	$\frac{1}{4}$	16.5	8	—	'repeat' too long
	13	$\frac{1}{2}$	—	—	—	no possible structure
	13	$\frac{1}{3}$	10.2	6	8	described in table 1
	13	$\frac{1}{4}$	≥ 10	8	—	'repeat' too long
	14	$\frac{1}{2}$	—	—	—	no possible structure
	14	$\frac{1}{3}$	10.2	6	9	described in table 1
	14	$\frac{1}{4}$	≥ 10	8	—	'repeat' too long
three-fold	10	$\frac{1}{2}$	13.4	6	—	'repeat' too long
	11	$\frac{1}{2}$	13.0	6	—	'repeat' too long

(i) *Seven-atom ring. One-half of CO and NH groups hydrogen bonded ($2_7 \cdot \frac{1}{2}$)*

In its fully extended state this structure has a repeat distance of 11.6 Å, with four amino-acid residues per repeat, and pairs of side-chains protruding alternately above and below the plane of the main chains in the *a* form. In the *b* form pairs of side-chains protrude alternately right and left within the plane of the main chain. Neighbouring side-chains are thus brought improbably close together. This proximity can be avoided in a variety of ways, however, by departing from the twofold screw symmetry. Such departure produces a slight shortening of the repeat distance. In end-on projection a structure of the type $S_7 \cdot \frac{1}{2}$ with high-order screw symmetry is

likely to have an appearance similar to that of 3_8 , 3_{10} or 4_{13} . The structure with twofold screw symmetry is illustrated in figure 13, which shows that neighbouring rings are free to rotate about the $C'-C-N$ bonds.

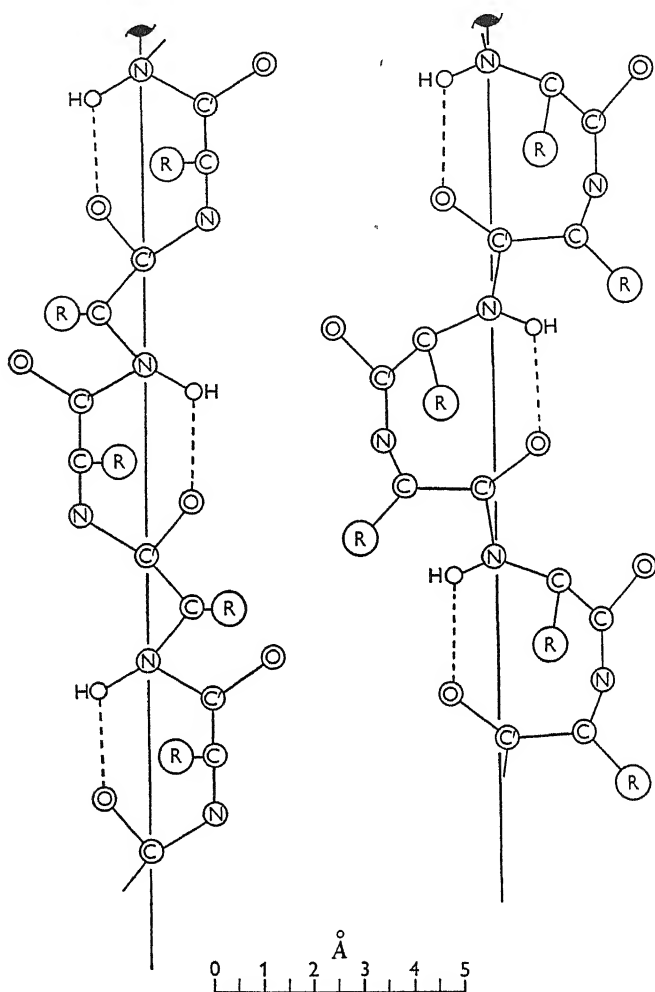


FIGURE 13

FIGURE 14

FIGURE 13. Sketch of the $2_{7.5}$ chain.

FIGURE 14. Sketch of the $2_{8.5}$ chain.

(ii) *Eight-atom ring. One-half of CO and NH groups hydrogen bonded ($2_{8.5}$)*

This structure has a repeat distance of 9.8 \AA with four amino-acid residues and is shown in figure 14. Neighbouring rings are free to rotate about the $N-C'$ bond, without appreciably affecting the length of the repeat. Side-chains protrude alternately parallel and perpendicular to the plane of the rings.

(iii) *Eleven-atom ring. One-third of NH and CO groups hydrogen bonded ($2_{11.3}$)*

This structure has a repeat distance of 9.6 \AA with six amino-acid residues, and is shown in figure 15. Side-chains protrude alternately up and down from the plane

of the rings. The properties of this model are very similar to those of 2_{13} and 2_{14} except that the individual rings are free to rotate about the C'—N bond. Viewed in end-on projection with the resolving power of the Pattersons of haemoglobin and myoglobin this structure would be indistinguishable from 2_{13} or 2_{14} .

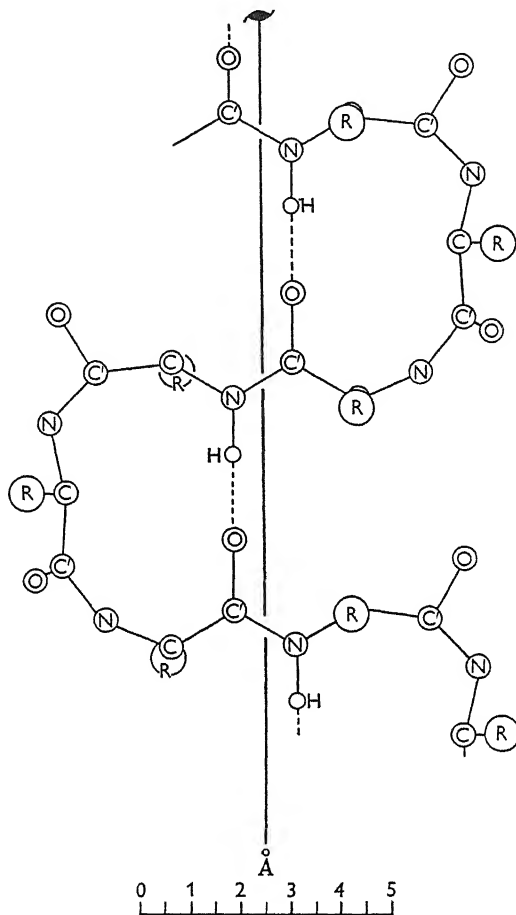


FIGURE 15. Sketch of the $2_{11.5}$ chain.

It is seen that in each of the three structures listed above the hydrogen bonded rings are free to rotate relative to each other. None of the models can be tied down to one specific configuration, so that definite parameters, on which a comparison of observed and calculated intensities might be based, cannot be assigned. Depending on the nature of the screw symmetry, end-on projections of these models would resemble one or the other of the types of structure described in detail in § 5. With the resolving power available no distinction can be made between models whose end-on projections are roughly similar.

(c) *Structures with secondary folds*

We may distinguish structures of a higher degree of complexity in which the 'asymmetric unit' which the screw axis repeats consists not of one hydrogen-

bridged ring alone, but of more than one ring or of some combination of rings and chain. Some of the structures in this category are similar to those described in the preceding paragraph and suffer from the same drawbacks. Examples of such structures are those recently discussed by Mizushima, Simanouti, Tsuboi, Sugita & Kato (1949), for example, these authors illustrate one in which the 'asymmetric unit' is formed of two 2_7b rings and one amino-acid residue which is not part of a ring; this asymmetric unit is operated upon by a screw dyad to form the chain.

Another structure of this type was proposed by Huggins (1943) and is illustrated in figure 16; here the asymmetric unit consists of three linked 2_7a rings, and the whole chain has screw dyad symmetry. Using space-filling models we have found the repeat distance of this chain to be 8 \AA , not 10.2 \AA as stated by Huggins. We ourselves have found it possible to coil a 2_7b structure into helical forms, with, for example, a screw hexad symmetry. The structure has an asymmetric unit consisting of a single seven-membered ring; the internal diameter of the helix is about 5 \AA and its pattern repeats at intervals of 6 \AA .

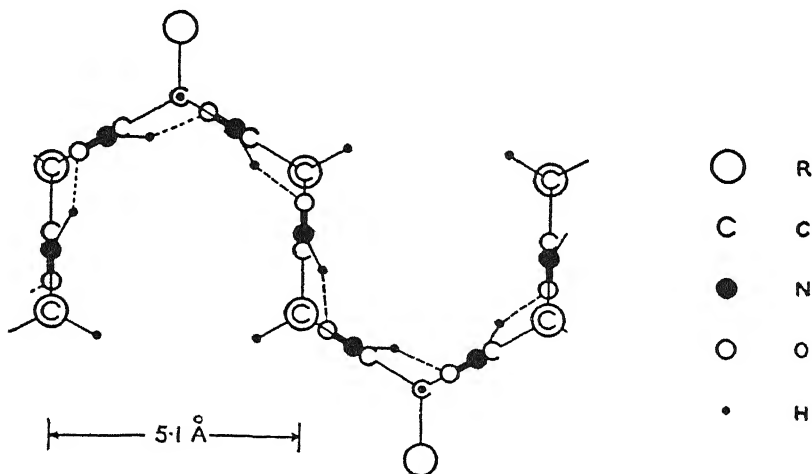


FIGURE 16. Structure with secondary folds built up of 2_7a rings, after Huggins (1943).

We have not made an exhaustive study of structures of this general type, because they all suffer from what we believe is a serious objection, namely, that, although the rings are held together by hydrogen bonds as in all the configurations mentioned earlier, the structure as a whole is not so held together. For example, there seems no obvious reason why Huggins's structure referred to in the last paragraph should not stretch to the ordinary 2_7b structure, since such stretching would involve only the overcoming of van der Waals attractions, and not the rupture of hydrogen bonds. It seems to be generally accepted that the α - β transformation in keratin involves the breaking of hydrogen bonds at least, though it must be admitted that up to now studies of the energetics of this transformation have not given any clear indication of the nature of the processes involved.

To sum up, after excluding the types of chain which are incompatible with three or four residues in a sub-cell of $10 \times 10 \times 5\text{ \AA}$, or are improbable for other reasons,

we are left with 2_7b , $2_{13}\cdot\frac{1}{3}$, $2_{14}\cdot\frac{1}{3}$, 3_8 , 3_{10} and 4_{13}^* . In the next section an attempt is made to narrow the possibilities still further by comparing the vector projections of these chains with the Patterson projections of haemoglobin and myoglobin. If, as is believed, the chains are parallel in their crystals, the vectors between atoms of the same chain should have a deciding influence on the form of the Patterson around the origin when the projection is upon a plane at right angles to the chain direction. This is a more favourable basis of comparison than the Patterson projections shown in figure 3 (*a*) and (*d*), because the chain vectors are crowded into a small area and their effects are superimposed.

7. VECTOR PROJECTIONS OF CHAIN MODELS

(a) End-on projections of the chains

A series of vector projections on a plane at right angles to the chain are shown in figure 17. The atoms C, N, O have been assigned equal weights, and vectors have been drawn between every pair of atoms in one repeat of the chain. These are transferred to a common origin, and their ends are plotted as projected on a plane. For instance, in the $2_{14}\cdot\frac{1}{3}$ chain which has six amino-acid residues in a repeat distance of $10\cdot2\text{ \AA}$ or 30 atoms per repeat, there are 900 vectors, 30 of which constitute the peak at the origin. The projection is divided into squares of 1 \AA side, and the number of vector ends in each square is counted. The density distribution can then be represented by contours at convenient levels. The vectors are distributed irregularly, and considerable smoothing is necessary in drawing the contours, but these contours serve to represent sufficiently accurately the type of Patterson diagram which the chains would give if arranged regularly in a simple unit cell with a wide spacing between the chains so that there is little overlap of the origin peaks. The correct representation would be attained by calculating the F 's of this arrangement, and constructing a Patterson projection with them. Such a calculation has been carried out for certain chains in the myoglobin cell in §8 below.

Figure 17 (*a*) is the vector projection of the $2_{14}\cdot\frac{1}{3}$ chain illustrated in figure 9. The atoms lie very nearly in a plane, with the exception of the first atoms R of the side-chains, and the Patterson projection is correspondingly elongated in one direction and narrow in the other. In comparing this projection with the origin peak of myoglobin, it must be remembered that the crystal has a dyad screw axis perpendicular to the chains, so that the part of the Patterson distribution around the origin due to the chains is a combination of (*a*) with its counterpart formed by rotation about the axis. Assuming the axis to be nearly parallel to the plane of the chain, the result would be as shown in figure 17 (*b*), which is formed by superposing (*a*) and its mirror image. Figure 17 (*e*) shows the distribution around the origin in a Patterson projection of myoglobin on a plane perpendicular to the chains, and it will be noted that it has a general similarity to (*b*). Figure 17 (*f*) is a similar projection for horse methaemoglobin. Its form suggests that it may be due to the superposition of

* The chains $2_7\cdot\frac{1}{2}$, $2_8\cdot\frac{1}{2}$ and $2_{11}\cdot\frac{1}{3}$ are also possible, but these chains are so ill-defined by the limiting conditions hitherto available that detailed tests of their conformity with the observations cannot be made at the present stage.

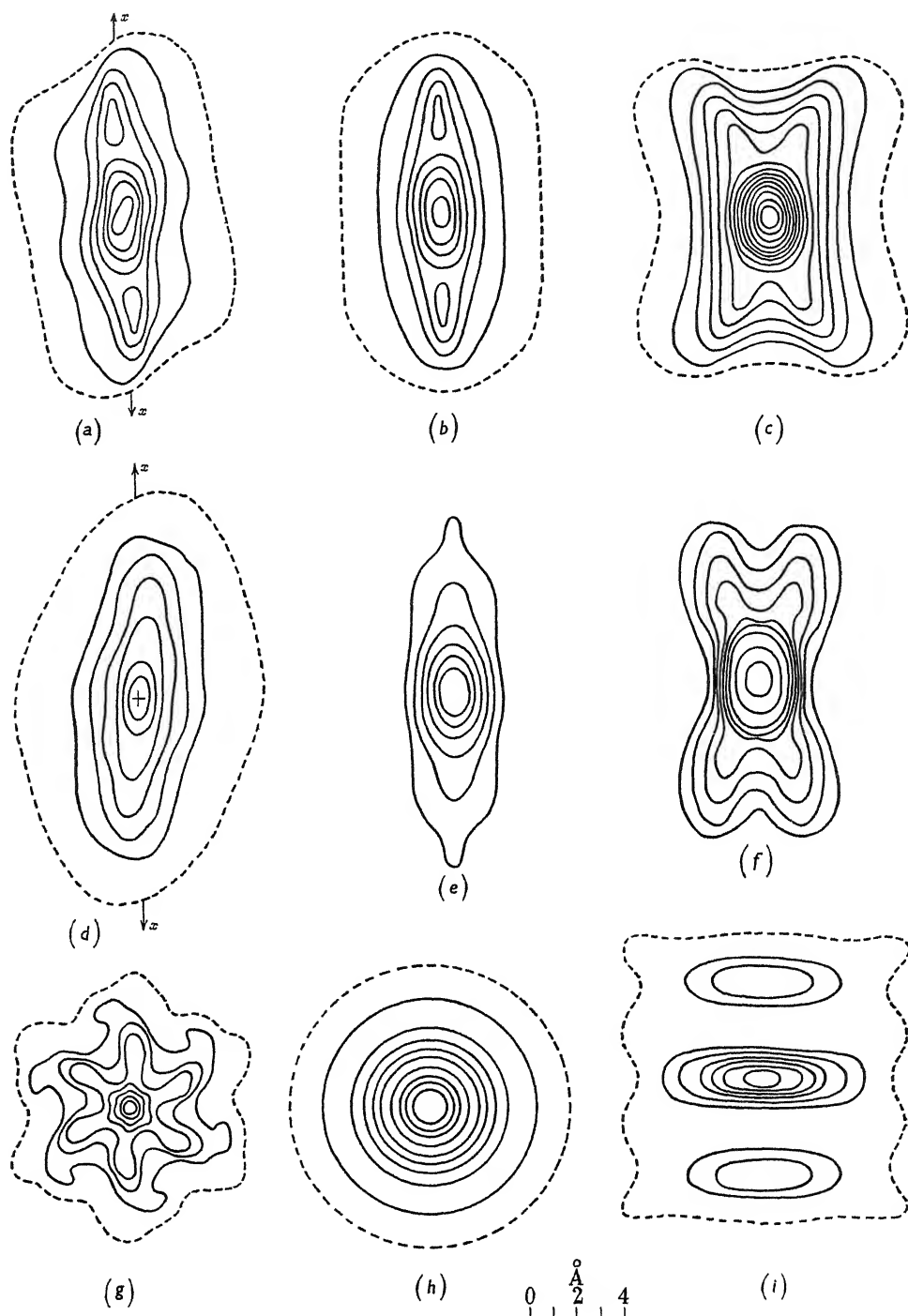


FIGURE 17. Origin peaks of vector projections along the chain direction. (a) $2_{14} \cdot \frac{1}{3}$ chain. (b) $2_{14} \cdot \frac{1}{3}$ chain, mirror image in plane parallel to plane of chain superimposed. (c) $2_{14} \cdot \frac{1}{3}$ chain; mirror image in plane inclined at 63° to plane of chain superimposed. (d) $2_{13} \cdot \frac{1}{3}$ chain. (e) Myoglobin: observed origin peak in $[20\bar{1}]$ projection. (f) Haemoglobin: observed origin peak in a projection. (g) 3_8 chain. (h) 3_8 chain, smoothed. (i) Pair of $2_7 b$ chains.

two projections such as (a) due to chains whose planes are inclined at an angle. Figure 17 (c) shows the combination of (a) with its counterpart about a twofold axis which makes an angle of 63° with the plane of the chain. There is a general similarity between (c) and (f). A general similarity is all that can be expected, since the Patterson projection of the crystal represents vectors between atoms of the side-groups, and between atoms of side-groups and chains, as well as intra-chain vectors. Conversely, the comparison between crystal Patterson and vector projection is not a very discriminating test for any one type of chain on account of this interference. Nevertheless, we can say that the observed Patterson projections are consistent with a type of folded chain such as $2_{14} \cdot \frac{1}{2}$.

Figure 17 (d) gives the vector projection of the $2_{13} \cdot \frac{1}{2}$ (Astbury) chain. Since the atoms of this chain diverge more from a coplanar arrangement than those of the $2_{14} \cdot \frac{1}{2}$ chain, the vector projection is less elongated and narrow. The two projections are so similar, however, that any comparison of the $2_{14} \cdot \frac{1}{2}$ chain with the observed peak applies equally to the $2_{13} \cdot \frac{1}{2}$ chain.

Figure 17 (g) is the vector projection of the screw triad chain 3_8 . If one takes into account the effect of the dyad symmetry of the haemoglobin crystal, and the smoothing resulting from the lack of resolution (least spacing, 2 \AA), one would expect the corresponding Patterson projection in a crystal to be more like figure 17 (h), where the density has been uniformly distributed so that it is a function of the radius alone. Any chain with a triad or tetrad axis might be expected to give a distribution which approximated to one of circular symmetry.

Figure 17 (i) is the vector projection of two parallel $2_7 b$ chains 4 \AA apart. It will be remembered that this is a possibility favoured by Ambrose *et al.* (1949).

At first sight, the comparison of the observed Patterson projections (e) and (f) with the vector projections of the various types of chain would appear to indicate rather definitely that the chain must be of the flat $2_{13} \cdot \frac{1}{2}$ or $2_{14} \cdot \frac{1}{2}$ type rather than types such as $2_7 b$, 3_8 , 3_{10} or 4_{13} . However, the very uncertain nature of the evidence has been pointed out to us by Mr F. H. C. Crick, who has drawn our attention to the point illustrated by figures 18 (a) and (b), which are based on the three-dimensional Patterson of haemoglobin. Figure 18 (a) is a projection of the Patterson density of a portion of the central rod (see figure 3 (a)) extending to 8.8 \AA on either side of the origin, including of course the central peak. It is not unlike figure 17 (f), suggesting two crossed vector projections of the (b) or (d) type. However, if a similar projection of the density is made for the portions between 7.8 and 23.4 \AA on either side of the origin, figure 18 (b) is the result. This is totally unlike figure 18 (a), being extended in the direction of the b axis and not at right angles to it. If the configuration of the main chain were the predominant factor in deciding the density in the rod, one would expect these two sections to be roughly similar, which is markedly not the case. Either the assumption that the chains possess a regular repeat of structure is incorrect, or, what is more probable, the irregular distortion due to atoms other than those in the main chains is so great as to mask the regular pattern due to the main chains alone. The comparisons illustrated by figure 17 are made with projections of the density of the complete unit cell, and so include all vectors from the shortest to the longest. We feel inclined to give more weight to the

complete projection or to figure 18 (*a*), both of which suggest a flat chain. It might be expected that the intra-chain vectors would stand out more clearly in the complete projection, since the disturbance due to the irregularly distributed atoms of the side-chains would be comparatively less when averaged over a greater volume. Again, if the chains were not straight or were in short lengths, the Patterson due to the shorter vectors (figure 18 (*a*)) should correspond more closely to the chain configuration. In any event, the comparison of figures 18 (*a*) and (*b*) shows how very tentative any conclusions must be; one can only say that the evidence favours the 'flat' chain such as $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$, but that it is inconclusive.

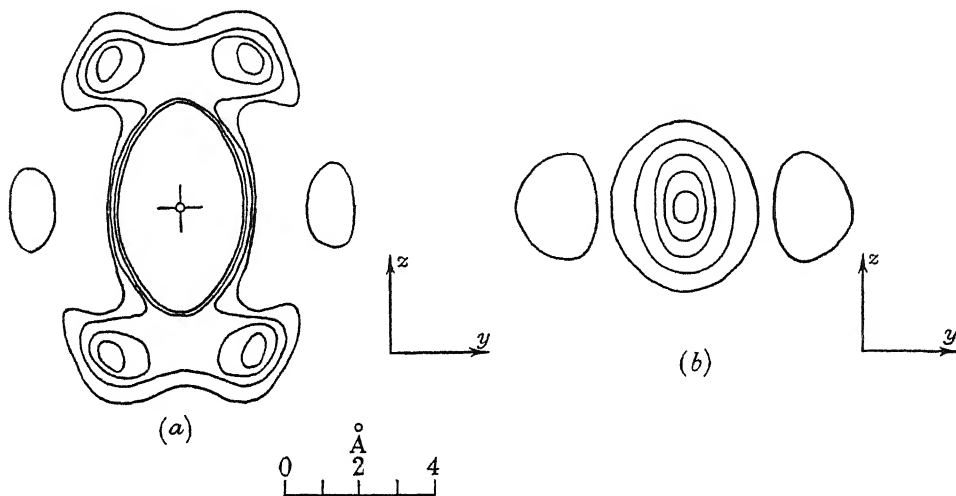


FIGURE 18. Projections along the *a* axis of portions of central rod of three-dimensional Patterson of haemoglobin. (*a*) Portion from $x = -8.8$ to $+8.8$ Å (*b*) Portions from $x = -23.4$ to -7.8 Å and $x = +7.8$ to $+23.4$ Å.

(*b*) *Lateral projections of the chains*

Vector projections on a plane parallel to the chains are shown in figure 19. Figure 19 (*a*) is a projection of the $2_{14} \cdot \frac{1}{3}$ chain on the *yz* plane. The atoms of this chain, apart from the R atoms, lie very nearly in the *xz* plane normal to the diagram. If they were exactly in this plane, all vectors would repeat at 5.1 Å intervals owing to the operation of the twofold screw axis parallel to *z*, and it will be noted that the vector diagram shows little difference between the origin peaks and the points midway between them. Figure 19 (*b*) is a projection of the same chain on the *xz* plane. In this case there is a strong concentration of vectors around the origin and a spreading at the mid-points, so that the 10.2 Å repeat is well marked. Figure 19 (*c*) is the *yz* projection of the $2_{13} \cdot \frac{1}{3}$ chain. Since the atoms of the chain depart more widely from the *xz* plane, there is a marked 10.2 Å repeat with a subsidiary 5.1 Å repeat. The 900 vectors in each case are spread over a wider area than in the 'end-on' projections, and it is consequently necessary to carry out a more drastic smoothing in attempting to indicate their density by contours, but these diagrams indicate the appearance of the Patterson projection to be expected in each case.

A lateral vector projection of the 3_8 chain contains only 225 points of which fifteen are at the origin, and it is spread over a wide area. It is difficult to draw

contours in such a case. In figure 19 (*d*), therefore, the projections of the terminations of the vectors are themselves plotted, in order to give an idea of their distribution. There is a bunching of fifteen vectors at the origin points, and a less marked concentration at a few other points, but apart from them the density is much the same for points along any line parallel to the axis. The Patterson, in other words, shows the 5.1 \AA repeat but no evidence of a spacing of one-third of this due to the threefold screw axis.

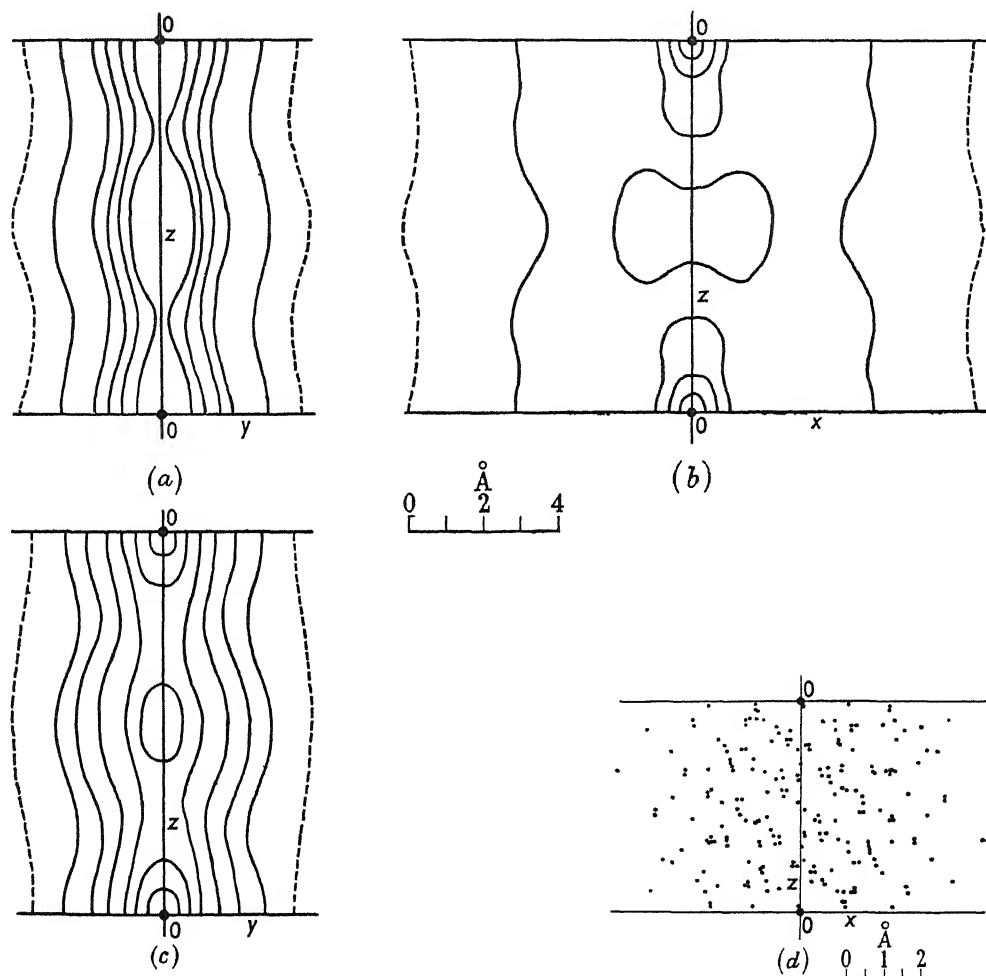


FIGURE 19. Lateral vector projections of chain models. (a) $2_{14} \cdot \frac{1}{3}$ chain, projection on to yz. (b) $2_{14} \cdot \frac{1}{3}$ chain, projection on to xz. (c) $2_{13} \cdot \frac{1}{3}$ chain, projection on to yz. (d) 3_3 chain, projection on to yz; only vector ends are shown, omitting 15 vectors at the origins.

8. INTRODUCTION OF VARIOUS TYPES OF CHAIN INTO THE MYOGLOBIN UNIT CELL (IN $[20\bar{1}]$ PROJECTION); COMPARISONS BETWEEN OBSERVED AND CALCULATED INTENSITIES

In a recent paper by one of us (Kendrew 1950) evidence has been given that the polypeptide chains in crystals of horse myoglobin lie parallel to the $[20\bar{1}]$ direction; this arrangement is shown in the vector projection along $[010]$, reproduced in

figure 3 (*d*) above. Accordingly a projection of the cell contents along $[20\bar{1}]$ should show the chains in end-on view; the Patterson diagram corresponding to this projection was calculated from the observed $|F|^2$'s of the $[20\bar{1}]$ zone and is reproduced below (figure 20 (*a*)). Consideration of the relative electron densities of 'main-chain' and 'side-chain' regions of the molecule showed that the principal features of such a projection were most probably due to the main chains; and the vector diagram was interpreted in terms of an arrangement of main chains of the kind illustrated in figure 21 (*a*) below. In this figure each pair of circles, joined by a line so

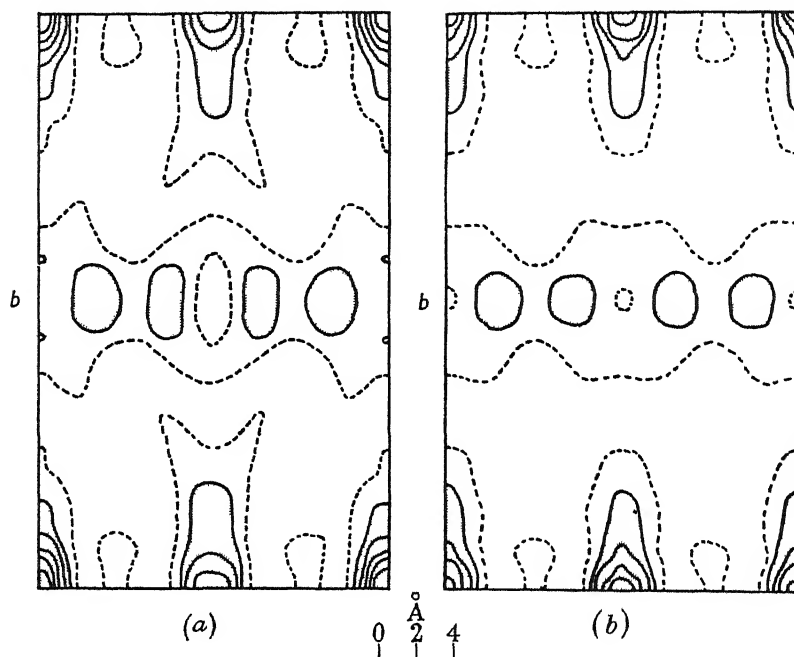


FIGURE 20. (*a*) $[20\bar{1}]$ Patterson projection of myoglobin. (*b*) $[20\bar{1}]$ Patterson projection of myoglobin, with terms for which $h=\text{odd}$ omitted; i.e. Patterson projection of actual structure with a glide plane superimposed.

that the centres of the circles are 4 \AA apart, might represent the end-on view either of one chain thrown into a series of planar folds parallel to b , or of a pair of chains 4 \AA apart (if not in three dimensions at least in projection). The molecule is believed to consist of a co-planar array of such chains (or pairs of chains) whose distance apart is equal to half the horizontal dimension of the projection, i.e. 8.47 \AA (see Kendrew 1950, figure 9 (*a*)); in neighbouring layers (or molecules) corresponding chains are 'staggered' by an amount twice the parameter p in figure 21 (*a*). The value of p , as measured from the Patterson projection, is in the neighbourhood of 0.070 (expressed as a fraction of the horizontal dimension of the projection).

We have attempted to test the validity of this scheme, and to compare the merits of various chain structures, by calculating the intensities of X-ray reflexions from a model projection made up of end-on projections of each type of chain in turn. These chains were placed in the two-dimensional unit cell in the way indicated in figure 21 (*a*). The intensities so calculated may be compared with those actually observed,

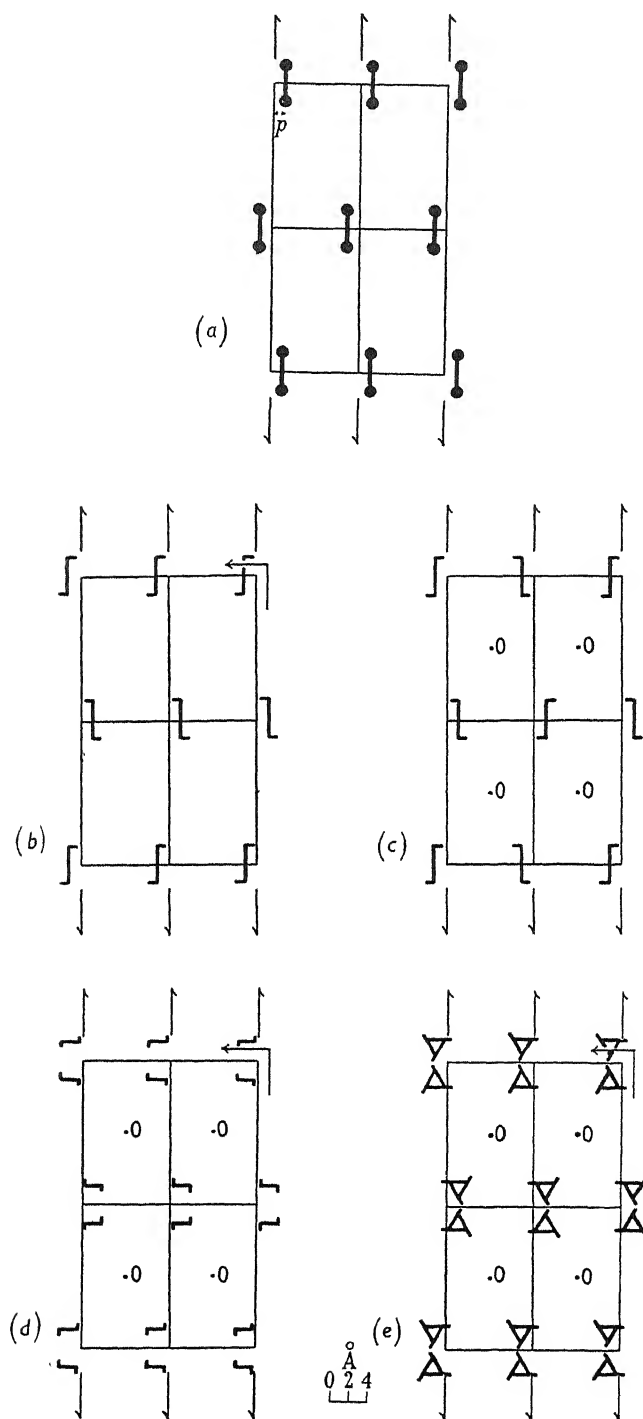


FIGURE 21. (a) Postulated arrangement of chains in myoglobin, as seen in $[20\bar{1}]$ projection. (b) Arrangement of $2_{13} \cdot \frac{1}{2}$ or $2_{14} \cdot \frac{1}{2}$ chains in myoglobin $[20\bar{1}]$ projection with glide plane. (c) As for (b), with centre of symmetry. (d) Arrangement of pairs of $2_7 b$ chains. (e) Arrangement of pairs of 3_{10} chains.

either directly or by calculating a Patterson projection from them. We have given reasons (in the preceding paragraph) why a model cell so constructed, and using the right chain configuration, should give a Patterson projection agreeing quite closely with that observed even though no account has been taken of the side chains (after the first carbon atom C_β), and hence comparisons between projection of trial cells using various types of chain should throw some light on the configuration actually present in myoglobin.

Briefly, the method of calculating the intensities of reflexions from the model cells was to compute a 'molecular scattering factor' for the end-on projection of a 'molecule' consisting of a single repeat unit of the chain. The 'molecular scattering factor' is exactly analogous to the atomic scattering factor of an atom; by combining it with an appropriate phase factor determined by the position of each 'molecule' in the cell, and summing over all the 'molecules', it is possible to obtain the amplitude and phase of any reflexion. The advantage of this procedure is that the parameters of the 'molecules' can be readily adjusted without recalculation of the molecular scattering factor; various arrangements of chains can thus be tried out with a minimum of computation.

The $2_{13} \cdot \frac{1}{3}$ (Astbury) and $2_{14} \cdot \frac{1}{3}$ chains

The end-on projections of the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ structures have been illustrated above (figures 8 and 9). Both projections have centres of symmetry, and there are two obvious ways in which such chains might be arranged in the cell; the first (figure 21 (b)) has a glide plane parallel to the plane of the paper (with translation perpendicular to b) and no centre of symmetry, while the second (figure 21 (c)) has a centre of symmetry but no glide plane parallel to the paper. Both, of course, have screw dyads parallel to $[010]$ in accordance with the space-group symmetry. However, in each type of chain the two arrangements give nearly identical values of F^2_{calc} , because the end-on projection of the chain itself approximates so closely to a higher symmetry (two mirror planes at right angles). Thus in effect the structure has both a centre of symmetry and a glide plane; hence $F(hk) = 0$ for odd h .*

Now, in the real crystal there is no glide plane perpendicular to $[20\bar{1}]$, in other words $F(hk)$ is not zero for odd h . However, reflexions for which h is odd are relatively weak (table 3), and it is not unreasonable to assume provisionally that they are due to the arrangement of side-chains, the main chains themselves being related by a glide plane. However, to make a direct comparison of the real with the hypothetical structures we must omit terms with odd h in comparing relative F^2 values or in computing Patterson projections. Figure 20 (b) shows the Patterson projection of the real structure calculated with these terms omitted (it can alternatively be regarded as being obtained from figure 20 (a) by reflecting the latter across mirror planes parallel to the (102) planes of the original cell at $x = \frac{1}{4}, \frac{3}{4}$).

Table 3 shows observed and calculated values of F^2 using the $2_{13} \cdot \frac{1}{3}$ (Astbury) chain inserted according to the scheme of figure 21 (b) or (c). The observed values

* For the remainder of this section we shall use x and y for the directions parallel to the sides of the $[20\bar{1}]$ projection, viz. the directions parallel to d_{102} and b respectively. h and k will be used to correspond, and $l=0$ throughout, e.g. the reflexion 142 is re-indexed as 140.

are only relative, so the calculated ones have been adjusted to give the same intensity for (020) as that observed.

These values of F^2_{calc} were used to calculate a Patterson projection which is shown in figure 22 (a).

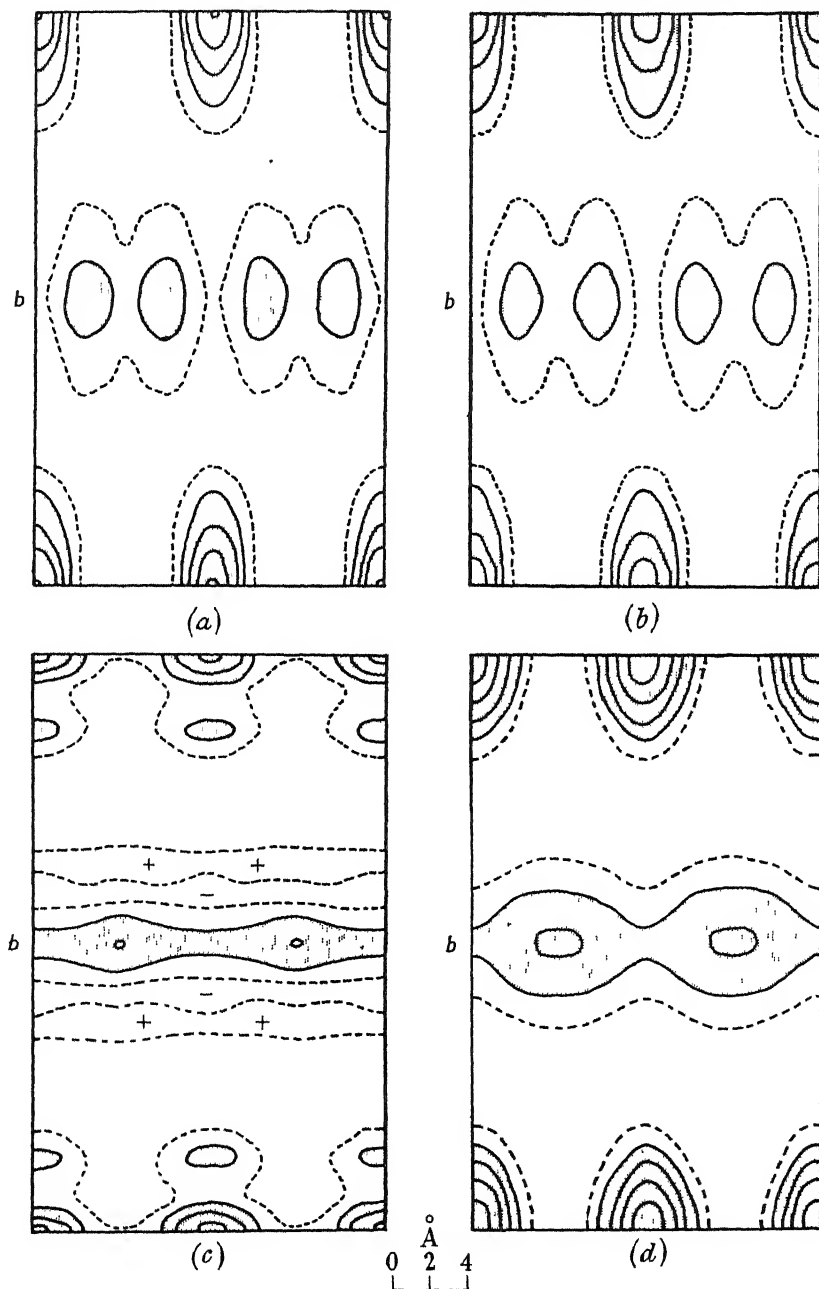


FIGURE 22. Model [20I] Patterson projection of myoglobin using (a) $2_{14} \cdot \frac{1}{3}$ chains, (b) $2_{13} \cdot \frac{1}{3}$ chains, (c) pairs of 2_7b chains, (d) pairs of 3_{10} chains.

It will be noted from table 3 that the model is reasonably successful in reproducing the four strongest reflexions in the pattern (020, 200, 210, 410), and a still better fit for 200 and 210 can be obtained by altering the parameter p slightly. The calculated Patterson projection (figure 22 (a)) resembles the observed projection (calculated from even h terms only, figure 20 (b)) remarkably closely. Besides the general arrangement of the peaks, which is in the main a consequence of the arrangement of chains in the cell rather than of the configuration of the chains themselves, the following points are noteworthy:

- (i) the shapes of the peaks in the $y = 0$ layer correspond closely with those observed though they are if anything too short and wide;
- (ii) the four peaks in the layer at $y = \frac{1}{2}$ are clearly resolved and closely resemble those in the observed projection.

Turning now to the $2_{14} \cdot \frac{1}{3}$ structure, the calculated values of F^2 are given in table 4 and the vector projection is given in figure 22 (b). The agreement between calculated and observed intensities is again reasonably close, and the vector projection is almost identical with that obtained from the $2_{13} \cdot \frac{1}{3}$ chain. Clearly it is impossible at this stage to discriminate between the two configurations.

The 2_7b chain

A prominent feature of the c projection of myoglobin is a pair of 4 \AA vectors parallel to b (Kendrew 1950, figure 4 (a)); in the $[20\bar{1}]$ projection too the main peaks show signs of containing three maxima 4 \AA apart in the b direction. These observations appeared on the face of it to fit well with the suggestion of Ambrose & Hanby (1949) that the 2_7b structure could fold back on itself about an axis parallel to the plane of the rings, study of models showed that the folds of the chain would lie about 4 \AA apart. The unit of structure would then consist of two chains 4 \AA apart running through the subcell, which would contain four residues.

To test the applicability of these ideas to myoglobin, the transform of the end-on projection of the 2_7b structure (figure 6) was calculated and used to calculate intensities of reflexions with an arrangement of chains in pairs as shown in figure 21 (d). The calculated intensities are compared with the observed values in table 5; it will be seen that the agreement is nothing like so good as in the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ structures. In particular, the model gives very strong 060 and 080 reflexions which are not observed in the real structure. The discrepancies are more obvious if the vector projection calculated from these values (figure 22 (c)) is compared with the observed projection (figure 20 (b)). The peaks in the $y = 0$ layer are altogether the wrong shape, and in the $y = \frac{1}{2}$ layer the four maxima are no longer resolved owing to the breadth of the projection of the chain-pair in the x direction.

The 3_{10} chain

The end-on projection of the 3_{10} helix (figure 11) is too large for two such chains to be placed 4 \AA apart. In this instance it is necessary to suppose that in the $[20\bar{1}]$ projection of myoglobin we are really viewing two molecules, one behind the other with chains displaced by 4 \AA relative to one another (see Kendrew 1950, p. 80).

Assuming an arrangement of this kind, pairs of chains have been inserted into the projection in the way indicated in figure 21 (*e*). The resulting values of F^2 are compared in table 6 with those observed, and it will be noticed again that the agreement is not so good as in the case of the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ configurations. The calculated vector projection (figure 22 (*d*)) fails in the following respects:

- (i) the origin peak is too thick and is insufficiently elongated; there is no sign of resolution into three peaks;
- (ii) in the middle layer the four peaks are unresolved. The same difficulties are encountered in the 3_8 and 4_{13} structures.

9. DISCUSSION OF RESULTS

If we base our conclusions only on the projections on a plane at right angles to the chain direction, undoubtedly the best agreement with the experimentally determined projections and intensities of diffraction is given by the $2_{13} \cdot \frac{1}{3}$ and $2_{14} \cdot \frac{1}{3}$ chains. Considering that the side-chains and intermolecular liquid have been left completely out of account, the resemblance between the projections (figures 22 (*a*), (*b*), and 20 (*b*)) is very striking. In contrast to this agreement, in the structures shown in figures 22 (*c*) and (*d*) the lack of resolution of peaks in the middle layer would appear to be a decisive criterion for rejecting these structures for myoglobin. The nearest peaks in the middle layer of the observed vector projection are only 4.2 Å apart; it can be shown by drawing cross-sections of the vector equivalent of the various chains (as illustrated in figure 17) that resolution could be expected only for chains about as narrow, in the x direction, as the $2_{13} \cdot \frac{1}{3}$ or $2_{14} \cdot \frac{1}{3}$ configurations.

The extension of the comparison to the three-dimensional Patterson projection of haemoglobin in §7, however, shows how misleading it may be to base conclusions on reflexions around a single zone. It will be necessary to get a three-dimensional Patterson for myoglobin in order to assess whether these resemblances are significant or fortuitous.

The problem is very complex, and one is forced to rely on a number of items of evidence each of which is very slight. The conclusion that the chains are of a folded coplanar form resembling the $2_{13} \cdot \frac{1}{3}$ or $2_{14} \cdot \frac{1}{3}$ type would be more convincing if there were any indication that the form has obvious advantages over others. In X-ray analysis in general, when a crystal structure has been successfully analyzed and a model of it is built, it presents so neat a solution of the requirements of packing and interplay of atomic forces that it carries conviction as to its essential correctness. In the present case the models to which we have been led have no obvious advantages over their alternatives. Much more evidence must be accumulated before conclusions can be safely drawn. There are certain hopeful features, however, at the present stage of the investigation. Assuming that criteria of the type adopted in this paper are justified, a survey shows that the number of possible forms of chain is very restricted. There appears to be a real simplicity of chain structure in myoglobin, which will perhaps be shown by other favourably built proteins, and which makes it particularly suitable for intensive X-ray investigation. There is hope that the study of such proteins may lead to a reliable determination of the chain structure.

The authors wish to thank Mr C. W. Bunn for his helpful advice on the technique of model building, Mr F. H. C. Crick for his valuable criticisms of a number of points raised in §7, and Miss V. E. Marting for calculating the Fourier transform of the 3_{10} helix. They also thank the Medical Research Council for financial support of two of them (J. C. K. and M.F.P.)

REFERENCES

- Albrecht, G. & Corey, R. B. 1939 *J. Amer. Chem. Soc.* **61**, 1087.
- Ambrose, E. J., Elliott, A. & Temple, R. B. 1949 *Nature*, **163**, 859.
- Ambrose, E. J. & Hanby, W. E. 1949 *Nature*, **163**, 483.
- Astbury, W. T. 1949*a* *Brit. J. Radiol.* **22**, no. 259.
- Astbury, W. T. 1949*b* *Nature*, **164**, 439.
- Astbury, W. T. & Bell, F. O. 1941 *Nature*, **147**, 696.
- Boyes-Watson, J., Davidson, E. & Perutz, M. F. 1947 *Proc. Roy. Soc. A*, **191**, 83.
- Bragg, W. L. 1949 *Nature*, **164**, 7.
- Chibnall, A. C. 1945 2nd Procter Memorial Lecture. Croydon: Int. Soc. of Leather Trades' Chemists.
- Corey, R. B. 1948 *Advances in protein chemistry*, **4**, 385. New York: Academic Press Inc.
- Darmon, S. E. & Sutherland, G. B. B. M. 1949 *Nature*, **164**, 440.
- Huggins, M. L. 1943 *Chem. Rev.* **32**, 195.
- Hughes, E. W. & Moore, W. J. 1949 *J. Amer. Chem. Soc.* **71**, 2618.
- Kendrew, J. C. 1950 *Proc. Roy. Soc. A*, **201**, 62.
- Kendrew, J. C. & Perutz, M. F. 1949 *Haemoglobin*, p. 161. London: Butterworths.
- Levy, H. A. & Corey, R. B. 1941 *J. Amer. Chem. Soc.* **63**, 2095.
- Mizushima, S., Simanouti, T., Tsuboi, M., Sugita, T. & Kato, E. 1949 *Nature*, **164**, 918.
- Perutz, M. F. 1949 *Proc. Roy. Soc. A*, **195**, 474.
- Sanger, F. 1948 *Nature*, **162**, 491.
- Taylor, H. S. 1941 *Proc. Amer. Phil. Soc.* **85**, 1.
- Zahn, H. 1947 *Z. Naturforschung*, **2b**, 104.

The decay of axisymmetric turbulence

BY S. CHANDRASEKHAR, F.R.S., *Yerkes Observatory*

(Received 1 April 1950)

In this paper the decay of axisymmetric turbulence is investigated. Explicit solutions, appropriate for the final period of decay, are obtained; these solutions are in agreement with Batchelor's general results on the role of the big eddies in homogeneous turbulence. It is further shown that, compatible with the equations which have been derived for axisymmetric turbulence, a state of turbulence exists which may be pictured as a superposition of two non-interacting fields which are isotropic and axisymmetric, respectively.

1. INTRODUCTION

The recent investigations of Batchelor (1948, 1949), Batchelor & Townsend (1948*a, b*) and Batchelor & Stewart (1950) have contributed greatly to our understanding of the processes underlying the decay of turbulence. In some ways the most remarkable result which is established in these papers is that tendency to isotropy does not exist among the large-scale components of homogeneous turbulence, and that the anisotropy resulting from this manifests itself most clearly during the last stages of the decay of turbulence. Thus, considering the Fourier transform

$$\Gamma_{ij}(\mathbf{k}) = \frac{1}{8\pi^3} \iiint Q_{ij}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r} \quad (1)$$

of the fundamental correlation tensor $Q_{ij}(\mathbf{r}) = \overline{u_i(0)u_j(\mathbf{r})}$, Batchelor (1949) has shown that as $|\mathbf{k}| = k \rightarrow 0$

$$\Gamma_{ij}(\mathbf{k}) = \Gamma_{ij;mn} k_m k_n + O(k^4), \quad (2)^*$$

where $\Gamma_{ij;mn}$ is a numerical fourth-order tensor which is constant during the decay; and, further, that during the last stages of the decay when viscosity is the determining factor,

$$\Gamma_{ij}(\mathbf{k}, t) \rightarrow \Gamma_{ij;mn} k_m k_n e^{-2\nu k^2(t-t_0)}. \quad (3)$$

The corresponding behaviour of the correlation tensor is

$$Q_{ij}(\mathbf{r}, t) = \frac{\frac{1}{8}\pi \sqrt{(\frac{1}{2}\pi)}}{[\nu(t-t_0)]^{\frac{1}{2}}} \left[\Gamma_{ij;mn} - \frac{\Gamma_{ij;mn} x_m x_n}{4\nu(t-t_0)} \right] e^{-r^2/8\nu(t-t_0)}. \quad (4)$$

Putting $r = 0$ in this last equation, we obtain the following expressions for the mean squares of the three velocity components:

$$\frac{\overline{u_1^2}}{\Gamma_{11;mm}} = \frac{\overline{u_2^2}}{\Gamma_{22;mm}} = \frac{\overline{u_3^2}}{\Gamma_{33;mm}} = \frac{\frac{1}{8}\pi \sqrt{(\frac{1}{2}\pi)}}{[\nu(t-t_0)]^{\frac{1}{2}}}. \quad (5)$$

These predictions of the theory have been confirmed by the experiments of Batchelor & Townsend (1948*b*) and Batchelor & Stewart (1950). While these investigations have clarified in a general way the role of the big eddies in determining the decay

* Here and elsewhere, summation over repeated indices is to be understood.

of turbulence in the final period, the question arises whether something more cannot be said about the manner in which the pattern of anisotropy (4) finally emerges. Since isotropy has been established not to prevail, it would be simplest to suppose that under the conditions of the experiments turbulence is *axisymmetric* (Batchelor 1946; Chandrasekhar 1950, referred to hereafter as I). And if the assumption of axisymmetry is made then, as we shall show, the equations of axisymmetric turbulence as developed in I enable us to trace in a definite manner how the final pattern of the turbulence emerges. This paper is principally devoted to this problem. However, the last section of the paper deals with a somewhat more general aspect of axisymmetric turbulence.

2. THE SOLUTION FOR THE SCALARS DEFINING Q_{ij} , WHEN THE INERTIAL TERMS IN THE EQUATIONS OF MOTION ARE NEGLECTED

In axisymmetric turbulence the fundamental correlation tensor $Q_{ij} = \overline{u_i u_j'}$ can be expressed in terms of two scalars Q_1 and Q_2 in the form (I, equations (48) to (50))

$$Q_{ij} = \text{curl} \left\{ Q_1 \epsilon_{ijk} \xi_k + Q_2 \lambda_j \epsilon_{ilm} \lambda_l \xi_m + \frac{1}{r} \frac{\partial Q_1}{\partial \mu} \xi_j \epsilon_{ilm} \lambda_l \xi_m \right\}, \quad (6)$$

where the various symbols have the same meanings as in I.

The equations governing Q_1 and Q_2 are (I, equations (118) to (120))

$$\frac{\partial Q_1}{\partial t} = 2\nu \Delta Q_1 + S_1$$

$$\text{and} \quad \frac{\partial Q_2}{\partial t} = 2\nu \Delta Q_2 + \frac{4\nu}{r^2} \frac{\partial^2 Q_1}{\partial \mu^2} + S_2, \quad (7)$$

$$\text{where} \quad \Delta = \frac{\partial^2}{\partial r^2} + \frac{4}{r} \frac{\partial}{\partial r} + \frac{1-\mu^2}{r^2} \frac{\partial^2}{\partial \mu^2} - \frac{4\mu}{r^2} \frac{\partial}{\partial \mu} \quad (8)$$

and S_1 and S_2 are two further scalars derived from the pressure and the inertial terms in the equations of motion (for their definitions see I, §10).

Now it is known (cf. Batchelor 1949; Batchelor & Townsend 1948*b*) that during the last stages of the decay of turbulence when viscosity is the controlling force we may neglect the inertial term in the equations of motion. Accordingly, under these same conditions we may neglect the terms S_1 and S_2 in equations (7); the equations for Q_1 and Q_2 then become

$$\frac{\partial Q_1}{\partial t} = 2\nu \Delta Q_1 \quad (9)$$

$$\text{and} \quad \frac{\partial Q_2}{\partial t} = 2\nu \Delta Q_2 + \frac{4\nu}{r^2} \frac{\partial^2 Q_1}{\partial \mu^2}. \quad (10)$$

2.1. The solution for $Q_1(r, \mu; t)$

Equation (9) is of the same form as the equation of heat conduction in an axially symmetric five-dimensional space. And we require to solve this equation with the boundary condition that at $t = 0$ (say) Q_1 is some given function of r and μ ; the

required solution can, therefore, be written down by the method of sources (cf., for example, Paterson 1941).

The solution of the equation of heat conduction in a five-dimensional space $(x_1, x_2, x_3, x_4, x_5)$, which corresponds to a unit source at $\mathbf{r} = \mathbf{r}'$ at $t = 0$ is

$$\frac{1}{(8\pi\nu t)^{\frac{5}{2}}} e^{-|\mathbf{r}-\mathbf{r}'|^2/8\nu t}. \quad (11)$$

The solution of equation (9) satisfying the boundary condition

$$Q_1 = Q_1(r, \mu; 0) \quad \text{at} \quad t = 0, \quad (12)$$

is, therefore,

$$Q_1(r, \mu; t) = \frac{1}{(8\pi\nu t)^{\frac{5}{2}}} \iiint \iiint e^{-|\mathbf{r}-\mathbf{r}'|^2/8\nu t} Q_1(r', \mu'; 0) dx'_1 dx'_2 dx'_3 dx'_4 dx'_5, \quad (13)$$

where the integration is to be carried out over the entire five-dimensional space.

Introducing polar co-ordinates

$$\begin{aligned} x_1 &= r \cos \theta, & x_2 &= r \sin \theta \cos \phi_1, & x_3 &= r \sin \theta \sin \phi_1 \cos \phi_2, \\ x_4 &= r \sin \theta \sin \phi_1 \sin \phi_2 \cos \phi_3, & x_5 &= r \sin \theta \sin \phi_1 \sin \phi_2 \sin \phi_3, \end{aligned} \quad (14)$$

and assuming that in (13) \mathbf{r} lies in the (x_1, x_2) -plane, we have

$$\begin{aligned} Q_1(r, \mu; t) &= \frac{1}{(8\pi\nu t)^{\frac{5}{2}}} \int_0^\infty \int_0^\pi \int_0^\pi \int_0^\pi \int_0^{2\pi} e^{-(r^2+r'^2-2rr'\cos\Theta)/8\nu t} \\ &\quad \times Q_1(r', \mu'; 0) r'^4 \sin^3 \theta' \sin^2 \phi'_1 \sin \phi'_2 dr' d\theta' d\phi'_1 d\phi'_2 d\phi'_3. \end{aligned} \quad (15)$$

In (15) we have written $\mu = \cos \theta$ and

$$\cos \Theta = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos \phi'_1. \quad (16)$$

The integrations over ϕ'_2 and ϕ'_3 in (15) are readily performed and we are left with

$$Q_1(r, \mu; t) = \frac{4\pi}{(8\pi\nu t)^{\frac{5}{2}}} \int_0^\infty \int_0^\pi \int_0^\pi e^{-(r^2+r'^2-2rr'\cos\Theta)/8\nu t} Q_1(r', \mu'; 0) r'^4 \sin^3 \theta' \sin^2 \phi'_1 dr' d\theta' d\phi'_1. \quad (17)$$

Now $e^{rr'\cos\Theta/4\nu t}$ in (17) can be expanded in terms of the Gegenbauer polynomials $C_m^{\frac{3}{2}}(\cos \Theta)$ (cf. Watson 1944, pp. 50 and 369) and the Bessel functions $I_{m+\frac{1}{2}}(rr'/4\nu t)$ of half-odd integral orders for the purely imaginary argument; thus

$$e^{rr'\cos\Theta/4\nu t} = 2^{\frac{3}{2}} \Gamma(\tfrac{3}{2}) \sum_{m=0}^{\infty} (m + \tfrac{3}{2}) C_m^{\frac{3}{2}}(\cos \Theta) \frac{I_{m+\frac{1}{2}}(rr'/4\nu t)}{(rr'/4\nu t)^{\frac{3}{2}}}. \quad (18)$$

Also, we have the relation (Watson 1944, p. 369)

$$\int_0^\pi C_m^{\frac{3}{2}}(\cos \Theta) \sin^2 \phi'_1 d\phi'_1 = \frac{2^2 m! [\Gamma(\tfrac{3}{2})]^2}{\Gamma(m+3)} C_m^{\frac{3}{2}}(\mu) C_m^{\frac{3}{2}}(\mu'). \quad (19)$$

Using these relations, we can reduce (17) to the form

$$\begin{aligned} Q_1(r, \mu; t) &= \frac{4\pi e^{-r^2/8\nu t}}{(8\pi\nu t)^{\frac{5}{2}}} \sum_{m=0}^{\infty} \frac{2^{\frac{3}{2}} [\Gamma(\tfrac{3}{2})]^3 (m + \tfrac{3}{2})}{(m+2)(m+1)} C_m^{\frac{3}{2}}(\mu) \\ &\quad \times \int_0^\infty \int_{-1}^{+1} e^{-r'^2/8\nu t} r'^4 (1-\mu'^2) Q_1(r', \mu'; 0) C_m^{\frac{3}{2}}(\mu') \frac{I_{m+\frac{1}{2}}(rr'/4\nu t)}{(rr'/4\nu t)^{\frac{3}{2}}} dr' d\mu'. \end{aligned} \quad (20)$$

Without loss of any essential generality, we may suppose that $Q_1(r, \mu; 0)$ can be expanded as a series in the Gegenbauer polynomials $C_n^{\frac{3}{2}}(\cos \Theta)$. And since Q_1 is required to be an even function of μ (cf. I, §6) the expansion will involve only the even-order polynomials $C_{2n}^{\frac{3}{2}}(\mu)$. Thus, let

$$Q_1(r, \mu; 0) = \sum_{n=0}^{\infty} q_{2n}^{(1)}(r) C_{2n}^{\frac{3}{2}}(\mu). \quad (21)$$

Inserting this expansion in (20) and using the orthogonality property (cf. Sommerfeld 1949, p. 232)

$$\int_{-1}^{+1} C_m^{\frac{3}{2}}(\mu) C_n^{\frac{3}{2}}(\mu) (1-\mu^2) d\mu = \delta_{mn} \frac{\pi(m+2)(m+1)}{2^2(m+\frac{3}{2})[\Gamma(\frac{3}{2})]^2}, \quad (22)$$

of the Gegenbauer polynomials, we obtain

$$Q_1(r, \mu; t) = \frac{e^{-r^2/8\nu t}}{32(\nu t)^{\frac{3}{2}}} \sum_{n=0}^{\infty} C_{2n}^{\frac{3}{2}}(\mu) \int_0^{\infty} e^{-r'^2/8\nu t} r'^4 q_{2n}^{(1)}(r') \frac{I_{2n+\frac{3}{2}}(rr'/4\nu t)}{(rr'/4\nu t)^{\frac{3}{2}}} dr'. \quad (23)$$

This is the required solution for Q_1 .

For $r = 0$ equation (23) becomes (cf. equation (26) below)

$$Q_1(0, t) = \frac{1}{48(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}} \int_0^{\infty} e^{-r^2/8\nu t} r^4 q_0^{(1)}(r) dr. \quad (24)$$

2.2. The solution for $Q_2(r, \mu; t)$

In terms of the solution (23) for Q_1 we can complete the solution of equation (10) for Q_2 , again by the method of sources. Thus

$$\begin{aligned} Q_2(r, \mu; t) = & \frac{e^{-r^2/8\nu t}}{32(\nu t)^{\frac{3}{2}}} \sum_{n=0}^{\infty} C_{2n}^{\frac{3}{2}}(\mu) \int_0^{\infty} e^{-r'^2/8\nu t} r'^4 q_{2n}^{(2)}(r') \frac{I_{2n+\frac{3}{2}}(rr'/4\nu t)}{(rr'/4\nu t)^{\frac{3}{2}}} dr' \\ & + 4\nu \int_0^t \frac{dt'}{[8\pi\nu(t-t')]^{\frac{3}{2}}} \iiint \left(\frac{1}{r^2} \frac{\partial^2 Q_1}{\partial \mu^2} \right)_{r', \mu', t'} e^{-|r-r'|^2/8\nu(t-t')} dx'_1 dx'_2 dx'_3 dx'_4 dx'_5, \end{aligned} \quad (25)$$

where $q_{2n}^{(2)}(r)$ is the coefficient of $C_{2n}^{\frac{3}{2}}(\mu)$ in the expansion of $Q_2(r, \mu; 0)$ in the even-order Gegenbauer polynomials.

3. THE ASYMPTOTIC FORM OF THE SOLUTION FOR $t \rightarrow \infty$

Since
$$\frac{I_{2n+\frac{3}{2}}(x)}{x^{\frac{3}{2}}} \rightarrow \frac{(\frac{1}{2})^{2n+\frac{3}{2}}}{\Gamma(2n+\frac{5}{2})} x^{2n} \quad (x \rightarrow 0), \quad (26)$$

it is evident that the successive terms of the series (23) decrease in the ratio $1/t^2$ as $t \rightarrow \infty$. Hence as $t \rightarrow \infty$,

$$Q_1(r, \mu; t) \rightarrow \frac{e^{-r^2/8\nu t}}{48(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}} \int_0^{\infty} e^{-r'^2/8\nu t} r'^4 q_0^{(1)}(r') dr', \quad (27)$$

or
$$Q_1(r, \mu; t) \rightarrow -\frac{\Lambda_1 e^{-r^2/8\nu t}}{48(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}}, \quad (28)$$

if
$$\Lambda_1 = -\int_0^{\infty} q_0^{(1)}(r) r^4 dr \quad (29)$$

is assumed to exist.

According to (27), Q_1 becomes independent of μ as $t \rightarrow \infty$. The corresponding behaviour of $Q_2(r, \mu; t)$ is, therefore,

$$Q_2(r, \mu; t) \rightarrow -\frac{\Lambda_2 e^{-r^2/8\nu t}}{48(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}}, \quad (30)$$

$$\text{if} \quad \Lambda_2 = -\int_0^\infty q_0^{(2)}(r) r^4 dr \quad (31)$$

is assumed to exist.

The mean squares of the velocities parallel and perpendicular to the direction of the mean flow are given by (I, equation (108))

$$\overline{u_{\parallel}^2} = -2Q_1(0, t) \quad \text{and} \quad \overline{u_{\perp}^2} = -(2Q_1 + Q_2)_{r=0}. \quad (32)$$

Hence, in the final period

$$\overline{u_{\parallel}^2} = \frac{\Lambda_1}{24(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}} \quad \text{and} \quad \overline{u_{\perp}^2} = \frac{2\Lambda_1 + \Lambda_2}{48(2\pi)^{\frac{1}{2}}(\nu t)^{\frac{3}{2}}}. \quad (33)$$

Though Q_1 and Q_2 become independent of μ as $t \rightarrow \infty$, the pattern of the turbulence which prevails is a non-trivial case of axisymmetric turbulence; for, according to I, equations (50), when Q_1 and Q_2 are functions of r and t only, the coefficients of the correlation tensor

$$Q_{ij} = A\xi_i\xi_j + B\delta_{ij} + C\lambda_i\lambda_j + D(\xi_i\lambda_j + \xi_j\lambda_i) \quad (34)$$

have the values

$$A = \frac{1}{r} \frac{\partial}{\partial r} (Q_1 + Q_2),$$

$$B = -\left(r \frac{\partial}{\partial r} + 2\right) Q_1 - \left[r(1 - \mu^2) \frac{\partial}{\partial r} + 1\right] Q_2,$$

$$C = \left(r \frac{\partial}{\partial r} + 1\right) Q_2,$$

$$\text{and} \quad D = -\mu \frac{\partial Q_2}{\partial r}. \quad (35)$$

Further, in this case the four scales of turbulence defined as in I, equations (112) and (117), become

$$L_A = q_1 + q_2, \quad L_B = q_1 + \mu^2 q_2, \quad L_C = 0 \quad \text{and} \quad L_D = -\mu q_2, \quad (36)$$

where

$$q_1 = -(\overline{u_{\parallel}^2} \overline{u_{\perp}^2})^{-\frac{1}{2}} \int_0^\infty Q_1(r, t) dr$$

and

$$q_2 = -(\overline{u_{\parallel}^2} \overline{u_{\perp}^2})^{-\frac{1}{2}} \int_0^\infty Q_2(r, t) dr. \quad (37)$$

For Q_1 and Q_2 given by equations (28) and (30)

$$\frac{q_1}{\sqrt{(\pi \nu t)}} = \sqrt{\frac{\Lambda_1}{2\Lambda_1 + \Lambda_2}} \quad \text{and} \quad \frac{q_2}{\sqrt{(\pi \nu t)}} = \frac{\Lambda_2}{\sqrt{\{\Lambda_1(2\Lambda_1 + \Lambda_2)\}}}. \quad (38)$$

It can be verified that the foregoing asymptotic forms of the solutions for $t \rightarrow \infty$ recover, in the framework of axisymmetric turbulence, Batchelor's result as expressed in equation (4).

4. A SPECIAL CASE OF AXISYMMETRIC TURBULENCE

From the equations (7) governing Q_1 and Q_2 it is apparent that if Q_1 is independent of μ then the solution for Q_2 is unaffected by Q_1 . And since according to the discussion of the preceding sections, Q_1 does become independent of μ during the last stages of the decay, it is not unlikely that the special case

$$Q_1 \equiv Q_1(r, t) \quad (39)$$

has some significance. In this case the scalar S_1 must also be independent of μ and the equations (7) become

$$\frac{\partial Q_1}{\partial t} = 2\nu \left(\frac{\partial^2}{\partial r^2} + \frac{4}{r} \frac{\partial}{\partial r} \right) Q_1 + S_1(r, t) \quad (40)$$

and
$$\frac{\partial Q_2}{\partial t} = 2\nu \Delta Q_2 + S_2(r, \mu, t). \quad (41)$$

A particularly simple example of the case (39) occurs when (cf. I, equations (95), (138), (139) and (148))

$$T_1 = T_2 = T_6 = 0, \quad T_4 \equiv T_4(r, t) \quad (42)$$

and $T_3 + \omega$ and T_5 are independent of r and μ ; for, in this case the scalars S_1 and S_2 have the values

$$S_1 = 2 \left(r \frac{\partial}{\partial r} + 5 \right) T_4 \quad \text{and} \quad S_2 = 0. \quad (43)$$

The corresponding form of the triple correlation tensor T_{ijk} is

$$\begin{aligned} T_{ijk} = & \frac{2}{r} \frac{\partial T_4}{\partial r} \xi_i \xi_j \xi_k - (\xi_i \delta_{jk} + \xi_j \delta_{ik}) \left(r \frac{\partial}{\partial r} + 3 \right) T_4 \\ & + 2T_4 \xi_k \delta_{ij} - 2(\lambda_i \delta_{jk} + \lambda_j \delta_{ik}) T_5 + 2T_3 \lambda_k \delta_{ij}. \end{aligned} \quad (44)$$

Returning to equations (40) and (41), we now have

$$\frac{\partial Q_1}{\partial t} = 2\nu \left(\frac{\partial^2}{\partial r^2} + \frac{4}{r} \frac{\partial}{\partial r} \right) Q_1 + 2 \left(r \frac{\partial}{\partial r} + 5 \right) T_4 \quad (45)$$

and
$$\frac{\partial Q_2}{\partial t} = 2\nu \Delta Q_2. \quad (46)$$

Equation (46) for Q_2 can be solved explicitly; we have (cf. equation (25))

$$Q_2(r, \mu; t) = \frac{e^{-r^2/8\nu t}}{32(\nu t)^{\frac{3}{2}}} \sum_{n=0}^{\infty} C_{2n}^{\frac{3}{2}}(\mu) \int_0^{\infty} e^{-r'^2/8\nu t} r'^4 q_{2n}^{(2)}(r') \frac{I_{2n+\frac{3}{2}}(rr'/4\nu t)}{(rr'/4\nu t)^{\frac{3}{2}}} dr'. \quad (47)$$

Thus, in this case the scalar Q_2 characteristic of the axisymmetric nature of the turbulence, evolves independently of Q_1 while Q_1 itself satisfies an equation identical in form with the Karman-Howarth equation in the theory of isotropic turbulence and admits the Loitsiansky invariant

$$\int_0^{\infty} Q_1(r, t) r^4 dr = \text{constant}. \quad (48)$$

We may therefore picture the turbulence as a superposition of two non-interacting fields which are isotropic and axisymmetric, respectively. It would be of interest to know whether this example of axisymmetric turbulence is merely a mathematical curiosity or has a deeper physical meaning.

REFERENCES

- Batchelor, G. K. 1946 *Proc. Roy. Soc. A*, **186**, 480.
 Batchelor, G. K. 1948 *Quart. Appl. Math.* **6**, 97.
 Batchelor, G. K. 1949 *Proc. Roy. Soc. A*, **195**, 513.
 Batchelor, G. K. & Stewart, R. 1950 *Quart. J. Mech. Appl. Math.* (in the Press).
 Batchelor, G. K. & Townsend, A. A. 1948a *Proc. Roy. Soc. A*, **193**, 539.
 Batchelor, G. K. & Townsend, A. A. 1948b *Proc. Roy. Soc. A*, **194**, 527.
 Chandrasekhar, S. 1950 *Phil. Trans. A*, **242**, 557.
 Paterson, S. 1941 *Phil. Mag. Ser. 7*, **32**, 384.
 Sommerfeld, A. 1949 *Partial differential equations of physics*. New York: Academic Press.
 Watson, G. N. 1944 *A treatise on the theory of Bessel functions*. Cambridge University Press.

Polarizability and internuclear distance in the hydrogen molecule and molecule-ion

BY R. P. BELL, F.R.S. AND D. A. LONG

Physical Chemistry Laboratory, University of Oxford

(Received 1 April 1950)

The polarizability of the hydrogen molecule-ion has been calculated for a range of internuclear distances, using various types of wave-functions. It is concluded that the values obtained for the polarizability are relatively insensitive to the choice of wave-function. Similar calculations have been made for the hydrogen molecule, using an LCAO wave-function. The value obtained for the polarizability at the equilibrium internuclear distance is in fair agreement with experiment and with the results of earlier calculations. The predicted variation of polarizability with internuclear distance agrees approximately with that deduced previously from the refractivities of hydrogen and deuterium, but is greater than that derived from the observed intensities of Rayleigh and Raman scattering.

INTRODUCTION

The most direct means of determining the variation of polarizability with internuclear distance is by observing the intensities of Raman lines relative to the intensity of the Rayleigh scattering. In a diatomic molecule there is only one internuclear distance ρ , and the three components of the polarizability are $\alpha_x = \alpha_y$ and α_z (the z -axis corresponding to the axis of the molecule) with invariants $\alpha = \frac{1}{3}(\alpha_z + 2\alpha_x)$ and $\gamma = \alpha_z - \alpha_x$. Writing $\alpha' = \partial\alpha/\partial\rho$, $\gamma' = \partial\gamma/\partial\rho$, the intensities are given by Placzek's theory as

$$\frac{I(\text{Raman})}{I(\text{Rayleigh})} = \frac{\hbar(\nu - \nu_k)^4}{8\pi^2 \mu \nu_k \nu^4 (1 - e^{-\hbar\nu_k/kT})} \frac{(\alpha'^2 + \frac{13}{45}\gamma'^2)}{(\alpha^2 + \frac{13}{45}\gamma^2)}, \quad (1)$$

where ν is the frequency of the incident light, $\nu - \nu_k$ that of the Raman line, and μ the reduced mass of the molecule. γ and γ' are obtainable from the degree of depolarization of the Rayleigh and Raman scattering respectively, so that α' can be calculated from the observed Raman-Rayleigh intensity ratio.†

The only measurements of this ratio for hydrogen are those of Bhagavantam (1931, 1932) whose latest value is 4.2×10^{-3} . Taking $\gamma/\alpha = 0.30$, $\gamma'/\alpha' = 1.22$ this leads to $\alpha' = 0.67 \times 10^{-16} \text{ cm}^2$. However, not much reliance can be placed upon this value, for two reasons. In the first place, equation (1) applies only when the directions of illumination and observation are mutually perpendicular, which was not so in Bhagavantam's measurements. In the second place, large errors are involved in the evaluation by photographic means of an intensity ratio of the order $1:10^3$, especially when the two lines involved have different wave-lengths. The observed ratio may therefore be considerably in error; in particular, it is likely to be too low, because the apparent intensity of the Rayleigh scattering may be increased by spurious scattering from dust or parts of the apparatus.

Bell (1942) has proposed an entirely different method for obtaining α' from the effect of isotopic substitution on the observed polarizabilities of molecules containing hydrogen. On account of anharmonicity, the average internuclear distance $\bar{\rho}$ will not be quite the same in hydrogen and deuterium compounds, and their polarizabilities should therefore also differ slightly, as is found to be the case experimentally. If α' can be assumed independent of ρ over the small range of distances involved, then $\alpha' = (\alpha_{\text{H}} - \alpha_{\text{D}})/(\bar{\rho}_{\text{H}} - \bar{\rho}_{\text{D}})$, where α_{H} and α_{D} are extrapolated to infinite wave-length to eliminate dispersion effects, and $\bar{\rho}_{\text{H}}$ and $\bar{\rho}_{\text{D}}$ can be evaluated from the spectroscopic anharmonicity constants. For the hydrogen molecule the observed refractivities of H_2 and D_2 lead to $\alpha' = 1.68 \times 10^{-16} \text{ cm}^2$ which is much higher than the value derived from Raman intensities. This method also shows that α' is positive, while equation (1) yields only its numerical value.

In view of the discrepancy between the two values of α' derived from experimental data we have investigated its calculation from first principles. Although a number of theoretical polarizability calculations have been carried out for the hydrogen molecule, the only one which treats the effect of internuclear distance is that of Hirschfelder (1935).‡ Using modifications of the Heitler-London electron-pair wave-function he obtained values in the range $\alpha' = 0.89 - 1.00 \times 10^{-16} \text{ cm}^2$, according to the parameters used. This result is nearer to the values from Raman intensities than to that from the isotope effect, but Hirschfelder's results for the equilibrium internuclear distance give a value for γ which is considerably too low (though the value of α is approximately correct). It was therefore decided to carry out further theoretical calculations with the molecular orbital (LCAO) type of function, which is in many respects more successful than the electron-pair treatment (Coulson 1937). Calculations were also made for the hydrogen molecule-ion H_2^+ , since this provides a simpler model for testing different types of function and a wide range of internuclear distances.

† Neither γ nor γ' is known with any accuracy, but this causes no great uncertainty in the value for α' .

‡ Mrowka (1933) has also reported results for varying internuclear distances, but, as will be shown later, his calculations are in error.

PRINCIPLES OF POLARIZABILITY CALCULATIONS

The available methods for calculating the polarizability can be classified as perturbation and variational. One form of the perturbation treatment involves expanding the perturbed wave-function in terms of the wave-functions of excited states of the unperturbed molecule (Mrowka 1933; Mulliken & Rieke 1941); this is a useful procedure when dispersion effects are being treated, but is a cumbersome method for stationary fields. In any case, there is some doubt about the accuracy of perturbation methods in general, since the second-order perturbation treatment sometimes gives results which are more in error than those of the first-order treatment, and we have therefore used the more powerful and flexible variation method.

Consider the hydrogen molecule, in which the nuclei are labelled a and b and the electrons 1 and 2. If Ψ_0 represents an approximate wave-function for the molecule in the absence of an electric field (which may itself be obtained by a variational treatment), the corresponding energy is given by

$$E_0 = \frac{\int \Psi_0 H_0 \Psi_0^* d\tau}{\int \Psi_0 \Psi_0^* d\tau}, \quad (2)$$

where
$$H_0 = -\nabla_1^2 - \nabla_2^2 - 2 \left(\frac{1}{r_{a1}} + \frac{1}{r_{b1}} + \frac{1}{r_{a2}} + \frac{1}{r_{b2}} \right) + \frac{2}{r_{12}} + \frac{2}{\rho}, \quad (3)$$

and (as in all later equations) energies and distances are measured in atomic units (a.u.) of $e^2/2a_0$ and a_0 respectively. If an electric field F is applied along the x -axis, the new Hamiltonian is

$$H' = H_0 - 2F(x_1 + x_2). \quad (4)$$

As variation functions for use with this Hamiltonian we may try either

$$\Psi' = \Psi_0 \{1 + X(x_1 + x_2)\} \quad (5)$$

or
$$\Psi'' = \Psi_0 \{1 + X'(x_1 + x_2) + X''(x_1 r_1 + x_2 r_2)\}, \quad (6)$$

where X , X' and X'' are parameters involving the field strength, and r_1 and r_2 represent the distances of the electrons from the centre of the molecule. Function (6), with two parameters, should be the more accurate, as it allows for the fact that the electrons become more polarizable at large distances. However, it is much more cumbersome to use, and we shall use only the one-parameter function (5) in this paper. This gives for the energy E' in presence of a field

$$E' = \frac{\int \Psi' H' \Psi'^* d\tau}{\int \Psi' \Psi'^* d\tau} = \frac{L - 4MF X + NX^2}{P + MX^2}, \quad (7)$$

where
$$L = \int \Psi_0 H_0 \Psi_0^* d\tau, \quad M = \int \Psi_0 (x_1 + x_2)^2 \Psi_0^* d\tau, \\ N = \int \Psi_0 (x_1 + x_2) H_0 (x_1 + x_2) \Psi_0^* d\tau, \quad P = \int \Psi_0 \Psi_0^* d\tau,$$

some of the terms vanishing by symmetry. For low field strengths this can be expanded in powers of F and X , giving

$$E' = E_0 - \frac{4M}{P} FX + \left(\frac{N}{P} + \frac{ML}{P^2} \right) X^2, \quad (8)$$

neglecting powers higher than the second. Minimization of E with respect to the parameter X gives

$$X = \frac{2MP}{NP - ML} F, \quad E' = E_0 - \frac{4M^2}{(NP - ML)} F^2, \quad (9)$$

and hence for the polarizability

$$\alpha_x = \frac{4M^2}{NP - ML}, \quad (10)$$

since the extra energy due to the field is $-\alpha_x F^2$.† Expressions exactly analogous to equations (4) to (10) will of course hold for polarization in the z -direction. The corresponding equations for the hydrogen molecule-ion are obtained by omitting throughout terms referring to the second electron.

THE HYDROGEN MOLECULE-ION

The only previous polarizability calculation for this molecule is that of Hirschfelder (1935), who used a form of wave-function due to Guillemin & Zener (1929),

$$\Psi_0 = e^{-Z'r_a} e^{-Z''r_b} + e^{-Z''r_a} e^{-Z'r_b}, \quad (11)$$

where Z' and Z'' are adjustable parameters. Using a one-parameter variation function (5) he calculated α_x and α_z for $\rho = 0.10, 1.00, 2.00, 3.00$ and 4.00 , which are too widely spaced to give a reliable estimate of α' at the equilibrium distance $\rho = 2.00$. His values are compared with our own in table 2.

Equation (11) contains two parameters, and a two-parameter equation giving a much more accurate energy value has been given by James (1935), i.e.

$$\Psi_0 = e^{-\delta\mu}(1 + c\nu^2), \quad (12)$$

where μ and ν are the confocal elliptic co-ordinates of the electron, and δ and c parameters. We have therefore used this function in our polarizability calculations. James (1935) reports an energy value for the equilibrium distance only, and we have first extended his energy calculations over a large range of distances. The results are given in table 1, which shows the values of the parameters δ and c and also the separate potential and kinetic energies, since it has been shown (Bell & Coulson 1945) that these constitute a more sensitive test of a wave-function than does the total energy. The wave equation for the hydrogen molecule-ion can also be solved exactly (Hylleraas 1931; Jaffé 1934), and the exact values of the total energy are given for comparison.

† The usual expression $-\frac{1}{2}\alpha_x F^2$ would give the energy in units of e^2/a_0 , while we are using units of $e^2/2a_0$. The resulting expression for the polarizability is of course in units of a_0^3 .

Table 1 shows that up to $\rho = 3.00$ a.u. the total energies calculated from the James function agree excellently with the exact values. The potential and kinetic energies at the equilibrium distance $\rho = 2.00$ a.u. are in close agreement with the virial theorem $T = -\frac{1}{2}V = -E$, and the values of T and V at other distances agree well with the curves derived by Bell & Coulson (1945). On the other hand, it is evident that the James function will fail for large values of ρ , and it is probably beginning to break down at $\rho = 5.00$ and 7.00 a.u. This is shown by the rapid rise in the parameter c , and by the fact that at $\rho = 7.00$ a.u. E is greater than the true limiting value -1.0000 , similarly, T and V do not appear to be tending to the correct limiting values of $+1.0000$ and -2.0000 respectively. However, it is clear that the function gives an accurate description of the molecule for values of up to at least 3 a.u.

TABLE 1. ENERGY OF THE HYDROGEN MOLECULE-ION (JAMES FUNCTION)

(Distance in units a_0 , energies in units $e^2/2a_0$)

ρ	δ	c	T	V	E		
					James	Hylleraas	Jaffé
0.50	0.448	0.0368	+2.8210	-2.2906	+0.5304	+0.5302	+0.5300
1.00	0.797	0.129	+1.9446	-2.8476	-0.9030	-0.9046	-0.9035
1.25	0.949	0.192	+1.6752	-2.7584	-1.0830	-1.0826	—
1.50	1.104	0.265	+1.4998	-2.6638	-1.1640	-1.1644	—
1.75	1.226	0.351	+1.3210	-2.5182	-1.1972	-1.1980	—
2.00	1.354	0.448	+1.2032	-2.4074	-1.2042	-1.2053	-1.2053
2.25	1.477	0.559	+1.1120	-2.3114	-1.1994	-1.1998	—
2.50	1.596	0.686	+1.0414	-2.2288	-1.1874	-1.1878	—
2.75	1.713	0.829	+0.9866	-2.1580	-1.1714	-1.1716	—
3.00	1.827	0.993	+0.9442	-2.0982	-1.1544	-1.1551	-1.1544
5.00	2.705	3.486	+0.8289	-1.8674	-1.0385	—	—
7.00	3.570	22.69	+0.8368	-1.8090	-0.9724	—	—
∞	—	—	(+1.0000)	(-2.0000)	(-1.0000)	-1.0000	-1.0000

The polarizability calculations were made with the one-parameter variation function (5), equations (7) to (10) being employed. All the integrals involved can be evaluated by standard methods, and only the results will be presented here. Table 2 gives the values obtained for α_x , α_z and α together with those of Hirschfelder (1935) for comparison. There is good agreement between our polarizability values and those of Hirschfelder, although the wave-function used by the latter author gives much less accurate values of the energy. This suggests that an even simpler function may give satisfactory values for the polarizability, and calculations were therefore also made with a LCAO function involving only one parameter.

This function, first used by Finkelstein & Horowitz (1928), can be written

$$\Psi_0 = e^{-cr_a} + e^{-cr_b}, \quad (13)$$

where c is an adjustable parameter, the best value of which for a given internuclear distance is found by the usual variation treatment. For present purposes it is more conveniently transformed to elliptic co-ordinates, giving

$$\Psi_0 = e^{-t(\mu+\nu)} + e^{-t(\mu-\nu)}, \quad (14)$$

where t is a new parameter defined by $t = c\rho$. The values of ρ corresponding to round values of t have been given by Coulson (1937), and these are used in table 3. The

TABLE 2. POLARIZABILITY OF THE HYDROGEN MOLECULE-ION

ρ (a.u.)	James function			Hirschfelder		
	α_x	α_z	α	α_x	α_z	α
0.10	—	—	—	0.04	0.04	0.04
0.50	0.062	0.070	0.065	—	—	—
1.00	0.110	0.165	0.128	0.11	0.17	0.13
1.25	0.140	0.247	0.176	—	—	—
1.50	0.169	0.356	0.232	—	—	—
1.75	0.211	0.540	0.321	—	—	—
2.00	0.250	0.749	0.417	0.25	0.72	0.40
2.25	0.292	1.050	0.546	—	—	—
2.50	0.334	1.445	0.704	—	—	—
2.75	0.375	1.897	0.882	—	—	—
3.00	0.416	2.688	1.173	0.41	2.67	1.16
4.00	—	—	—	0.54	7.95	3.0
5.00	0.609	—	—	—	—	—
7.00	0.481	—	—	—	—	—

polarizabilities were calculated as before using a one-parameter variation function. The values given for α_x , α_z and α at $\rho = 0$ and for α_x at $\rho = \infty$ are obtained by the corresponding variational treatment of He^+ and He (Hassé 1930, 1931).

TABLE 3. POLARIZABILITIES OF THE HYDROGEN MOLECULE-ION—
LCAO FUNCTION

(Units of 10^{-24} cm^3)				
t (a.u.)	ρ (a.u.)	α_x	α_z	α
0	0	0.0371	0.0371	0.0371
0.5	0.2623	0.0456	0.0483	0.0465
1.0	0.5745	0.0702	0.0850	0.0751
1.5	0.9666	0.112	0.160	0.128
2.0	1.4556	0.186	0.413	0.262
2.5	2.0293	0.291	0.975	0.519
3.0	2.6448	0.406	2.092	0.968
4.0	3.8600	0.577	7.299	2.818
5.0	4.9898	0.696	19.11	6.835
7.0	7.0000	0.621	—	—
10.0	10.0000	0.594	—	—
∞	∞	0.593	∞	—

Figures 1 and 2 show the variation of α_x and α_z with ρ for the two types of wave-function. The data for the equilibrium separation $\rho = 2.00$ are given in table 4.

TABLE 4. POLARIZABILITY DATA FOR H_2^+ —EQUILIBRIUM
INTERNUCLEAR DISTANCE

	α_x	α_z	α	α_x	α'_z	α'
	10^{-24} cm^3			10^{-16} cm^2		
James function	0.25	0.75	0.42	0.31	1.90	0.83
LCAO function	0.29	0.94	0.51	0.34	2.02	0.91
Guillemin-Zener function	0.25	0.72	0.40	—	—	—

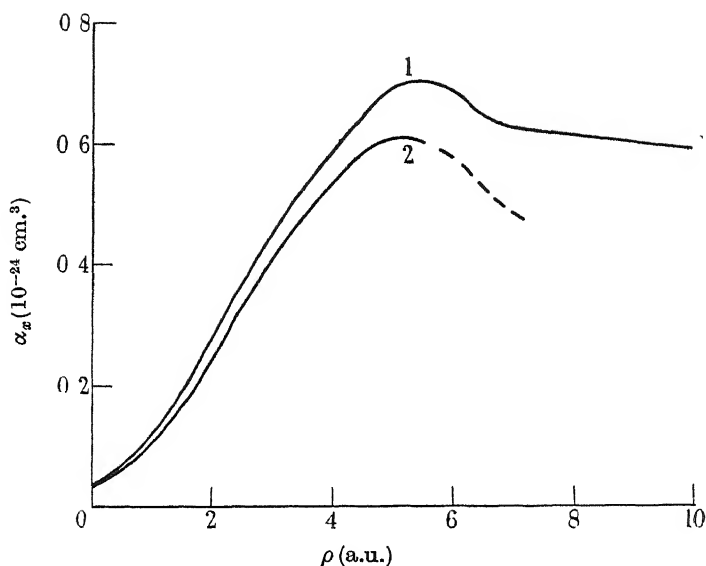


FIGURE 1. Curve 1, LCAO function; curve 2, James function.

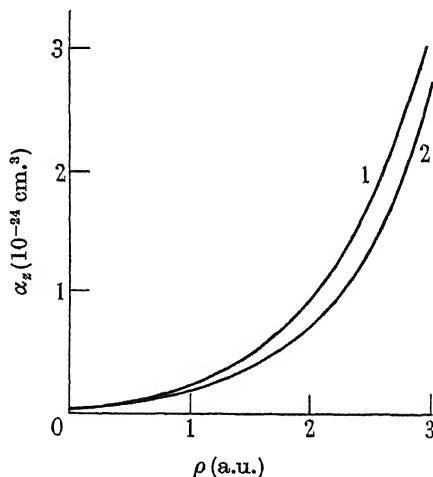


FIGURE 2. Curve 1, LCAO function; curve 2, James function.

DISCUSSION OF RESULTS FOR THE HYDROGEN MOLECULE-ION

The main result of the above calculations is that the values of α and α' are not very sensitive to the choice of wave-function for the unperturbed molecule, the LCAO function with one parameter giving results which are close to those derived from the accurate James function with two parameters. This suggests that the LCAO treatment may give reliable results with other systems, including the hydrogen molecule.

The numerical values of α and α' cannot be compared with experiment, though they are of the same order as those observed for the hydrogen molecule. However, the shapes of the curves in figures 1 and 2 are of some interest. The variation of

α_x with ρ (figure 1) follows the general course predicted by Bell (1942) for H_2 , except in the region $\rho = 4$ to 7 a.u., where it rises slightly above the limiting value for a hydrogen atom. It is not certain whether this is a real effect, or whether it is due to the inadequacy of the LCAO treatment. The curve derived from the James function shows signs of the same behaviour, but we have already seen that this function becomes inaccurate for the unperturbed molecule at large internuclear distances.

On the other hand, in both treatments the axial polarizability α_z tends to infinity as ρ increases, though it might be expected to approach the value for a hydrogen atom. However, this result is not surprising when considered more closely. Any acceptable wave-function for H_2^+ represents (in the absence of an external field) an electron distribution which is symmetrical about the centre of the molecule, and this remains true for any value of ρ . When ρ becomes great the electron is almost certain to be found in the neighbourhood of one nucleus or the other, both regions having equal probabilities. When an electric field is applied along the axis of the molecule these probabilities are no longer equal, and the energy difference between the two states $H^+ \dots H$ and $H \dots H^+$ in the field will increase without limit as the internuclear distance increases, corresponding to the infinite limiting value found for α_z . However, this result only has a meaning for observations in a stationary field, and as ρ increases the time necessary for a significant observation will also increase without limit. For large values of ρ it will be possible initially to locate the electron in the neighbourhood of one or other of the nuclei, and an average distribution symmetrical with respect to the two nuclei will be observed only if the time of observation is much greater than $h/\Delta E$, where ΔE is the value of the exchange integral between the two states. For H_2^+ , $h/\Delta E \sim 10^{-14}$ seconds for $\rho = 5$ a.u., so that if measurements were made with light of ordinary frequencies, dispersion effects would set in at about this internuclear distance, and the observed polarizability would fall rapidly to that of the hydrogen atom for greater values of ρ .

THE HYDROGEN MOLECULE

Table 5 contains a summary of previous theoretical treatments of the polarizability of the hydrogen molecule, mostly at the equilibrium distance only. Unless otherwise stated, the energy in the field was calculated by a variational method with one parameter. Of the wave-functions listed in table 5, (a) is the simple Heitler-London function composed of $1s$ atomic functions. Analogous functions are used for the excited states occurring in the treatments of Mrowka and Mulliken. (b) resembles (a), but with an adjustable screening constant in the exponents of the atomic functions as suggested by Wang (1928). The difference between the results obtained by Hirschfelder and by Steensholt with this function is due to their choice of slightly different screening constants. (c) is the modification of the Wang function suggested by Rosen (1931), in which the $1s$ functions are replaced by a linear combination of $1s$ and $2p_z$ functions, thus introducing a second adjustable constant. (d), (e) and (f) are the molecular orbital functions in elliptic co-ordinates proposed by Coulson (1938), involving respectively 2, 3 and 5 adjustable constants.

The values given by Mrowka are anomalous in that they make $\alpha_x > \alpha_z$. A later note by Mrowka (1938) indicated a correction making $\alpha_x < \alpha_z$, but no new values were given, and Mrowka's work has been criticized by Easthope (1936) and by Mulliken & Rieke (1941), whose calculations by the same method give $\alpha_x < \alpha_z$ in agreement with the variation treatments and with experiment. Omitting Mrowka's values, all the remaining calculations give a polarizability α in reasonable agreement with the experimental value of $0.79 \times 10^{-24} \text{ cm}^3$; this shows again that the polarizability is not very sensitive to the wave-function chosen for the unperturbed molecule. Further, the results obtained by Hirschfelder show that little is gained by including a second parameter to represent the effect of the electric field, as in equation (6).

TABLE 5. POLARIZABILITY OF THE HYDROGEN MOLECULE
(α in 10^{-24} cm^3 , α' in 10^{-18} cm^2)

author	wave- function	α_x	α_z	α	γ	α'_x	α'_z	α'
Hirschfelder (1935)	(c)	0.67	0.71	0.69	0.04	0.80	1.08	0.89
Hirschfelder (1935)	(b)	0.66	0.71	0.68	0.05	0.89	1.02	0.93
Hirschfelder (1935) (2 parameters)	(c)	0.74	0.75	0.74	0.01	0.95	1.10	1.00
Hirschfelder (1935) (2 parameters)	(b)	0.73	0.75	0.74	0.02	0.98	1.02	0.99
Steensholt (1935)	(b)	0.52	0.77	0.60	0.25	—	—	—
Steensholt (1947)	(d)	0.66	0.68	0.67	0.02	—	—	—
Steensholt (1947)	(e)	0.64	0.71	0.66	0.07	—	—	—
Steensholt (1947)	(f)	0.62	0.71	0.65	0.09	—	—	—
Easthope (1936) (perturbation)	(a)	0.59	0.72	0.62	0.13	—	—	—
Mrowka (1933) (excited states)	(a)	0.85	0.61	0.77	-0.24	-0.24	+0.42	0.20
Mulliken & Rieke (1941) (excited states)	(a)	0.70	0.92	0.77	0.22	—	—	—

It is more difficult to compare the calculated values of the anisotropy γ with experiment, since there is little agreement between the observed values of different authors for the degree of depolarization of the Rayleigh scattering; the measurements of Cabannes (1928), Parthasarathy (1932) and Volkmann (1935) correspond respectively to $10^{24}\gamma = 0.36, 0.35$ and 0.20 . The low values of γ given by most of the theoretical treatments are probably incorrect, but the errors involved in α_x and α_z are small.

We have calculated the polarizability of the hydrogen molecule using for the unperturbed molecule the LCAO function

$$\Psi_0 = \{\psi_a(1) + \psi_b(1)\}\{\psi_a(2) + \psi_b(2)\}, \quad (15)$$

where

$$\psi_a(1) = e^{-cra_1}, \quad \text{etc.}, \quad (16)$$

c being an adjustable parameter depending on ρ . It is again convenient to introduce a parameter defined by $t = c\rho$, and we have calculated polarizabilities for $\rho = 1.0, 1.5$ and 2.0 a.u. When (15) is inserted in (3), (7) and (10) the resulting expression for the polarizability contains a large number of different integrals. All those not involving r_{12} can be evaluated in a straightforward though tedious manner by transforming to elliptic co-ordinates, when they reduce to functions of $I_n(t)$ and $I_n(-t)$ ($n = 0, 1, 2, 3$ or 4), where

$$I_n(t) = \int_1^\infty \mu^n e^{-t\mu} d\mu. \quad (17)$$

Some of the appropriate numerical values of $I_n(t)$ have been tabulated by Rosen (1931). The integrals involving r_{12} offer greater difficulty, and only about one-quarter of them have been previously evaluated (Rosen 1931; Kemble & Zener 1929, Hirschfelder & Linnett 1950). The evaluation of the remainder will be described in a separate publication.

The results finally obtained for the polarizability are given in table 6. The values of ρ corresponding to round values of t were kindly supplied by Professor C. A. Coulson. The limiting values for $\rho = 0$ and $\rho = \infty$ are from one-parameter variation treatments of the systems He and 2H (Hassé 1930, 1931)

TABLE 6. POLARIZABILITIES OF THE HYDROGEN MOLECULE—
LCAO FUNCTION
(Units of 10^{-24} cm.³)

t (a.u.)	ρ (a.u.)	α'_x	α'_z	α
0	0	0.150	0.150	0.150
1.000	0.680	0.189	0.273	0.217
1.500	1.217	0.555	0.672	0.594
1.644†	1.375	0.66‡	0.83‡	0.72‡
2.000	1.811	0.953	1.28	1.06
∞	∞	1.19	∞	—

At $\rho = 1.375$, $\alpha'_x = 1.31$, $\alpha'_z = 1.86$, $\alpha' = 1.49$ ($\times 10^{-16}$ cm.²).

† Equilibrium distance.

‡ Interpolated values.

DISCUSSION OF RESULTS FOR THE HYDROGEN MOLECULE

Comparison of tables 5 and 6 shows that the LCAO treatment predicts an equilibrium mean polarizability which is closer to the experimental value 0.79×10^{-24} cm.³ than that predicted by any of the other wave-functions, when the effect of the field is represented by one parameter. This is true even for functions like (c), (d), (e) and (f) which involve 2 to 5 adjustable constants, compared with one only in equation (15). Moreover, the use of two field parameters with the Rosen and Wang functions produced a result which is little better than the LCAO result with one field parameter. In addition to this the LCAO value $\gamma = 0.17 \times 10^{-24}$ cm.³ is nearer the uncertain experimental result $\gamma = 0.2 - 0.4 \times 10^{-24}$ cm.³ than most of the values in table 5, so that the LCAO treatment appears to give a good picture of the polarizability of the hydrogen molecule in its equilibrium state.

The LCAO treatment predicts a derived polarizability $\alpha' = 1.49 \times 10^{-16}$ cm.², which is considerably nearer to the value 1.68×10^{-16} deduced from the isotope effect than to the much lower value 0.67×10^{-16} from Raman intensities. Moreover, table 6 shows that α' varies little with ρ over a small range in the neighbourhood of the equilibrium position, thus confirming the assumption on which the treatment of the isotope effect is based.

It should be noted that at $\rho = 2.0$ a.u. the calculated value of α_z is already somewhat greater than the expected limiting value for separation into two hydrogen atoms, and it is evident that the calculated value would increase without limit with increasing internuclear distance. This behaviour is reminiscent of that found for the

hydrogen molecule-ion, but is here due to the inadequacy of the LCAO treatment for large values of ρ . The function (15) gives equal weight to the covalent state $H + H$ and to the two ionic states $H^+ + H^-$ and $H^- + H^+$, and it is well known that this gives rise to error for large internuclear separations. In the absence of an electric field the two ionic states are equally probable, but if an external field is applied along the z -axis one will be more probable than the other, and the molecule will acquire an induced dipole moment which increases without limit as ρ increases. This difficulty does not arise for α_x (applied field perpendicular to the molecular axis), nor should it occur with a wave-function which automatically allows for the decrease in the importance of the ionic states as ρ increases. It is possible that this spurious increase in α_z may cause the value of α' given by the LCAO treatment to be a little too high. It is doubtful, however, whether the error is serious near the equilibrium internuclear distance, since the value predicted for α is below the experimental value, and Coulson (1937) has shown that the vibration frequency of the hydrogen molecule is predicted accurately by the LCAO treatment.

Our best thanks are due to Professor C. A. Coulson, Mr J. W. Linnett and Mr J. S. de Wet for advice and assistance in the above calculations, and to the Department of Scientific and Industrial Research for a grant to one of us (D.A.L.).

REFERENCES

- Bell, R. P. 1942 *Trans. Faraday Soc.* **38**, 422.
 Bell, R. P. & Coulson, C. A. 1945 *Trans. Faraday Soc.* **41**, 141.
 Bhagavantam, S. 1931 *Indian J. Phys.* **5**, 319, 557.
 Bhagavantam, S. 1932 *Indian J. Phys.* **7**, 107, 549.
 Cabannes, J. 1928 *J. Phys. Radium*, **4**, 929.
 Coulson, C. A. 1937 *Trans. Faraday Soc.* **37**, 1479.
 Coulson, C. A. 1938 *Proc. Camb. Phil. Soc.* **34**, 204.
 Easthope, C. E. 1936 *Proc. Camb. Phil. Soc.* **32**, 260.
 Finkelstein, B. N. & Horowitz, G. E. 1928 *Z. Phys.* **48**, 118.
 Guillemin, V. & Zener, C. 1929 *Phys. Rev.* **34**, 999.
 Hassé, H. R. 1930 *Proc. Camb. Phil. Soc.* **26**, 542.
 Hassé, H. R. 1931 *Proc. Camb. Phil. Soc.* **27**, 66.
 Hirschfelder, J. O. 1935 *J. Chem. Phys.* **3**, 555.
 Hirschfelder, J. O. & Linnett, J. W. 1950 *J. Chem. Phys.* **18**, 130.
 Hylleraas, E. A. 1931 *Z. Phys.* **71**, 739.
 Jaffé, G. 1934 *Z. Phys.* **87**, 535.
 James, H. M. 1935 *J. Chem. Phys.* **3**, 9.
 Kemble, E. C. & Zener, C. 1929 *Phys. Rev.* **33**, 512.
 Mrowka, B. 1933 *Z. Phys.* **84**, 448.
 Mrowka, B. 1938 *Eucken-Wolf Hand- und Jahrbuch d. Chem. Physik*, **10** (iii), 25.
 Mulliken, R. S. & Rieke, C. 1941 *Rep. Progr. Phys.* **8**, 231.
 Parthasarathy, S. 1932 *Indian J. Phys.* **6**, 139.
 Rosen, N. 1931 *Phys. Rev.* **38**, 255, 2099.
 Steensholt, G. 1935 *Z. Phys.* **93**, 620.
 Steensholt, G. 1947 *Phil. Mag.* (7) **38**, 748.
 Volkmann, H. 1935 *Ann. Phys., Lpz.*, **24**, 457.
 Wang, S. C. 1928 *Phys. Rev.* **31**, 579.

The nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K

I. Thermodynamic measurements made from a study of the field-dependence of the adiabatic susceptibility

BY C. G. B. GARRETT, PH.D.

The Royal Society Mond Laboratory, University of Cambridge

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 29 March 1950)

Thermodynamic measurements have been made at temperatures below 1° K, obtained by the method of magnetic cooling, on copper potassium sulphate and on a diluted copper Tutton salt. A study has been made of the field-dependence (for small fields) of the adiabatic susceptibility of the cooled and thermally isolated salt, the measurements covering the range of temperature from 1° K down to 0.05° K for copper potassium sulphate, and to 0.025° K for the dilute salt. From these measurements the entropy and magnetic susceptibility are determined as functions of the absolute temperature. It is concluded that for both salts the susceptibility follows a Curie-Weiss law, the values of Δ being 0.034 and 0.0048° K respectively; the specific heats are of the form A/T^2 , the values found for A being 6.1×10^{-4} R for copper potassium sulphate and 1.98×10^{-4} R for the dilute salt. Deviations from this behaviour in a ferromagnetic direction are found for copper potassium sulphate below 0.07° K.

1. INTRODUCTION

In addition to the use of magnetic cooling experiments simply for the attainment of very low temperatures, information may be obtained from them about the thermal and magnetic properties of the paramagnetic salts used in the magnetic cooling technique. The first object of such experiments is to determine the magnetic susceptibility and entropy of the material as a function of absolute temperature. Since both of these quantities will depend upon the spectrum of the low-lying energy levels of the paramagnetic ions, the experimental results are of interest in connexion with the quantum theory of paramagnetism. In this paper and part II (Garrett 1950*b*) results are presented of magnetic cooling experiments on copper potassium sulphate and on a sample of the same salt diluted about seven times with diamagnetic magnesium, from which it is possible to deduce the magnitude of that part of the specific heat which is due to nuclear-electronic interaction (Gorter 1948; Garrett 1949). This paper is concerned with the principle of the method used in the measurements of specific heat and absolute temperatures, and with the experimental results. In part II these results are related to the theory of interionic interactions, and the nuclear specific heat is derived.

In order to determine absolute temperatures and specific heats at temperatures below 1° K it is necessary to combine the results of two different sets of experiments. One set of results is obtained from a quantitative study of the demagnetizations themselves, the magnetic susceptibility χ at some low temperature being related to the initial field and initial temperature from which the demagnetization was carried out. Since the entropy of the initial state may be calculated by the standard methods of

statistical thermodynamics, a series of such experiments will determine the relation between the entropy S and the susceptibility of the final states of the material, provided that the demagnetizations can be assumed to be isentropic.

Another set of experimental results is now required before absolute temperatures and specific heats can be deduced. In the present investigation this has been provided by superimposing on the small alternating measuring field a steady parallel field H , and determining the decrease in susceptibility so produced. It will be shown in §2 that, apart from a small saturation correction term, the proportionate change in susceptibility is $\frac{3}{2}\Xi H^2$, where

$$\Xi = \frac{1}{\chi} \frac{d\chi}{dT} \frac{d\chi}{dS}.$$

It is clear that a combination of measurements of Ξ and of S as functions of χ makes possible the determination of the absolute temperature (see §3), and conversely S and χ become determined as functions of T .

To give a physical picture of this second type of experiment it is instructive to compare it with the bridge method of Casimir & du Pré (1938) for the investigation of paramagnetic relaxation, in which the experimental technique is exactly the same, except that the sample of salt is immersed in a bath of liquid helium. At high frequencies the quantity measured is the adiabatic susceptibility $(\partial M/\partial H)_S$, which will be smaller than the isothermal susceptibility $(\partial M/\partial H)_T$ measured at low frequencies; in some intermediate frequency range the susceptibility is complex, the real part lying between these two limits. In the experiments described in this paper, however, the conditions differ in two respects from those of the Casimir-du Pré experiment. In the first place the lattice thermal capacity is usually very small in comparison with that of the spin system, so that the adiabatic susceptibility is observed at all frequencies. The entire spin-lattice relaxation effect has disappeared (Casimir 1940). Further, the sample of salt is thermally isolated, instead of being immersed in a heat bath; and owing to the magneto-caloric effect the temperature is higher than it was in zero field, so that the susceptibility observed is even lower than the adiabatic value at the original temperature. The observed decrease in differential susceptibility is therefore due partly to the magneto-caloric rise in temperature produced by the steady field, and partly to the field-dependence, at any one temperature, of the adiabatic susceptibility.

2. THERMODYNAMICS OF THE FIELD-DEPENDENCE OF ADIABATIC SUSCEPTIBILITY

In discussing the thermodynamics of a paramagnetic salt it is convenient to refer extensive quantities to one gram-ion of the material, and throughout this paper entropy, specific heat and susceptibility are quoted in this way. In a reversible change in which the temperature changes from T to $T + dT$ and the field from H to $H + dH$, the heat dQ taken in is given according to the second law of thermodynamics by the equation

$$dQ = C_H dT + T \left(\frac{\partial M}{\partial T} \right)_H dH, \quad (1)$$

where M is the intensity of magnetization and C_H the specific heat at constant field strength. To determine the adiabatic susceptibility, equation (1) is rewritten with H and M as the independent variables, dQ being equated to zero. It is found that

$$\left(\frac{\partial M}{\partial H}\right)_{S:(T',H)} = \left(\frac{\partial M}{\partial H}\right)_{T:(T',H)} \left\{ 1 - \frac{T}{C_H} \left[\left(\frac{\partial M}{\partial T}\right)_H \right]^2 \left(\frac{\partial H}{\partial M}\right)_T \right\}. \quad (2)$$

In the experiment, an isolated sample of salt is initially at temperature T and in zero field, the adiabatic and isothermal susceptibilities being then identical. A steady field H is now applied; equation (2) gives the adiabatic susceptibility, but only in terms of the isothermal susceptibility at the new temperature T' , which, as explained in § 1, is somewhat higher than T . The difference in temperature may be derived from equation (1):

$$\Delta T = - \int_0^H \frac{T}{C_H} \left(\frac{\partial M}{\partial T}\right)_H dH, \quad (3)$$

so that the isothermal susceptibility at the new temperature is

$$\left(\frac{\partial M}{\partial H}\right)_{T:(T',H)} = \left(\frac{\partial M}{\partial H}\right)_{T:(T,H)} - \frac{\partial^2 M}{\partial H \partial T} \int_0^H \frac{T}{C_H} \left(\frac{\partial M}{\partial T}\right)_H dH + \dots \quad (4)$$

In order to allow for the onset of saturation effects it is convenient to expand M as a power series in H :

$$M = \chi H + \psi H^3 + \dots, \quad (5)$$

where χ, ψ, \dots are functions of the temperature only. Substituting in equations (2) and (4) it is found that

$$\left(\frac{\partial M}{\partial H}\right)_{S:(T',H)} \simeq \chi \left[1 - \left(\frac{3}{2} \Xi - \frac{3\psi}{\chi} \right) H^2 \right], \quad (6)$$

ignoring terms in higher powers of H . In equation (6)

$$\Xi = \left(\frac{d\chi}{dT}\right)^2 \frac{T}{\chi C_H} = \frac{1}{\chi} \frac{d\chi}{dT} \frac{d\chi}{dS}, \quad (7)$$

since C_H differs from $T(dS/dT)$ only by terms of order H^2 . The term $3\psi/\chi$ which appears in equation (6) is a saturation correction and is small in comparison with $\frac{3}{2}\Xi$ throughout the accessible range of temperature for both the salts studied in this paper. The value of $3\psi/\chi$ at each temperature may be calculated from statistical thermodynamical considerations, details of which appear in the appendix.

3. USE OF THE Ξ FUNCTION

At sufficiently high temperatures it is usually a good approximation to write $C = A/T^2$ (ignoring for the moment the lattice specific heat) and $\chi = \lambda/T$ (Curie's law). Under these conditions Ξ is easily seen from equation (7) to have the temperature-independent value λ/A . This behaviour is well shown by potassium chrome alum (Garrett 1948) at temperatures above 0.3°K ; A may be estimated from the temperature-independent value. At lower temperatures deviations will occur both from the Curie law and from the A/T^2 specific heat law; generally these deviations are in such a sense as to make Ξ increase considerably above the value λ/A .

So long as the Curie law is valid, the absolute temperature may be derived directly from the measured value of χ , being given simply by λ/χ . At lower temperatures it is often convenient to define an arbitrary 'magnetic' temperature $T^* = \lambda/\chi$, which will of course differ more or less seriously from the true thermodynamic temperature. In this lower temperature range the true temperature may be derived from a knowledge of Ξ and of $dS/d\chi$ as functions of χ (and so of T^*); for it follows from equation (7) that

$$T = \int^{\chi} \frac{d\chi}{\chi \Xi(dS/d\chi)} = - \int^{T^*} \frac{dT^*}{T^* \Xi(dS/d\chi)}, \quad (8)$$

and, as all the quantities under the integral sign are known, the temperature T may be derived graphically by integration. One fixed point is necessary in order to specify the lower limit in the integration.

A useful special case is that in which the A/T^2 specific law is followed with sufficient accuracy, while the susceptibility follows a Curie-Weiss law:

$$\chi = \frac{\Lambda}{T - \Delta}. \quad (9)$$

Under these conditions it follows from equation (7) that

$$\Xi = \frac{\Lambda}{A} \left(\frac{T}{T - \Delta} \right)^3 = \frac{\Lambda}{A} \left(1 + \frac{\chi \Delta}{\Lambda} \right)^3. \quad (10)$$

In an actual experiment only relative susceptibility measurements are made. When Δ is of the order of 0.01° it is impossible to detect departures from the Curie law ($\chi T = \text{constant}$) in the helium range. Below 1°K , T^* is then defined by

$$T^* = \chi_c T_c / \chi, \quad (11)$$

where χ_c refers to a representative calibration temperature T_c , which is usually taken to be 2.5°K . With this definition of T^* it follows that, so long as the Curie-Weiss law is true,

$$T = T^*[(T_c - \Delta)/T_c] + \Delta. \quad (12)$$

In the region of the A/T^2 specific heat law it follows that

$$S = -\frac{A}{2T^2} = -\frac{1}{2}A \left[T^* \frac{T_c - \Delta}{T_c} + \Delta \right]^{-2}. \quad (13)$$

When it is suspected that the Curie-Weiss and A/T^2 laws are good approximations, the experimental data are analyzed in the following way:

(i) A graph is plotted of $\Xi^{\frac{1}{3}}$ against χ . This should be a straight line (equation (10)), from the gradient and intercept of which A and Δ may be found.

(ii) Using this value of Δ and the demagnetization data, a graph is plotted of S against $[T^*(T_c - \Delta)/T_c + \Delta]^{-2}$. This graph should also be straight (equation (13)), and the gradient ($\frac{1}{2}A$) provides an independent estimate of the specific heat constant A .

(iii) The breakdown of the two laws is indicated by departures from linearity in graphs (i) and (ii). At lower temperatures the absolute temperature may now be derived from the general formula (equation (8)), taking as fixed point some convenient temperature on the linear portions of the two graphs.

At and above 0.5°K these considerations break down, because the lattice specific heat (BT^3) begins to be appreciable in comparison with that of the spin system. Since the ratio of these two quantities is proportional to T^5 , the influence of the lattice will be detected as a sudden decline in the measured values of Ξ , provided that the frequency of the measuring field is sufficiently low. Near the onset of the lattice decline it is a reasonable approximation to put $\Xi = \lambda/(A + BT^5)$, so that it should in principle be possible to deduce from the measurements the lattice specific heat constant B . In practice it was found that the estimates of B so made were larger than was reasonable on other grounds, probably the discrepancy is due to the extra thermal capacity of small quantities of occluded helium.

In carrying out magnetic cooling experiments it is customary to have the salt in spheroidal shape, and subsequently to correct measurements of susceptibility and of T^* to the definitive shape of the sphere by means of the formulae

$$\chi_{\text{sphere}} = \frac{\chi_{\text{spheroid}}}{1 + \left(\frac{4\pi}{3} - \alpha\right) \frac{\chi_{\text{spheroid}}}{V}}, \quad (14)$$

$$T_{\text{sphere}}^* = \frac{T_{\text{spheroid}}^* \left[1 + \left(\frac{4\pi}{3} - \alpha\right) \frac{\chi_{\text{spheroid}}}{V} \right]}{1 + \left(\frac{4\pi}{3} - \alpha\right) \frac{(\chi_c)_{\text{spheroid}}}{V}} \\ \simeq T_{\text{spheroid}}^* + \left(\frac{4\pi}{3} - \alpha\right) \frac{(T^*\chi)_{\text{spheroid}}}{V}, \quad (15)$$

in which V is the gram-ionic volume and α is the demagnetizing factor of the spheroid. Since at a typical calibration temperature $\left(\frac{4\pi}{3} - \alpha\right) \frac{(\chi_c)_{\text{spheroid}}}{V}$ is generally of the order of 0.001° or less, the error introduced by the approximation in equation (15) is $\sim \frac{1}{10}\%$ for all values of T^* . In addition, it is necessary to correct the observed values of Ξ to spherical shape by the relation

$$\Xi_{\text{sphere}} = \Xi_{\text{spheroid}} \left\{ 1 + \left(\frac{4\pi}{3} - \alpha\right) \frac{\chi_{\text{spheroid}}}{V} \right\}^{-3}, \quad (16)$$

whose validity is obvious from the definition of Ξ (equation (7)). It will be noticed that equation (8) gives the same value of the absolute temperature whether or not a shape correction is applied. For a compressed powder α may be calculated by means of an interpolation formula (Casimir 1940)

$$\alpha = \alpha_e + \frac{4\pi}{3} \frac{1-f}{f}, \quad (17)$$

where f is the filling factor of the sample and α_e is the demagnetizing factor appropriate to its external shape. In this paper all values of χ , T^* and Ξ have been corrected to spherical shape, and no special notation is used to draw attention to this fact.

4. EXPERIMENTAL DETAILS

The technique of magnetic cooling experiments has now become fairly standardized and requires no detailed description here. The magnet used was a water-cooled armoured solenoid of the type developed by Ashmead, developing a maximum field of about 10 kG. It was mounted on a supporting ram which could be raised or lowered when required by means of a worm drive from a small motor.

The salt sample consisted of a polycrystalline powder compressed to a filling factor of about 0.95, ground on a lathe to spheroidal shape. It was suspended inside the german silver salt-tube of the demagnetization cryostat by means of fine nylon or cotton threads from a german silver framework.

Measurements of magnetic susceptibility were made by determining the mutual inductance between two coils magnetically linked with the salt, the primary having about 400 turns over a length of 10 cm., and the secondary having 5000 turns over a length of 5 cm. Compensating secondary coils, each of 2500 turns, were included in order to reduce the mutual inductance of the empty coil system to a low value. The mutual inductance was measured by means of a Hartshorn bridge (Casimir 1940) using a c. at a frequency of 40 c./sec. from a small frequency-stabilized rotary converter. The bridge output was detected by means of a vibration galvanometer preceded by a tuned amplifier. With this arrangement the mutual inductance due to the salt was usually of the order of several thousand μH ; for a peak measuring field of 0.3 G the mutual inductance could be set to 0.1 μH with the detector at low sensitivity and short setting time, or to 0.01 μH at the cost of an increased setting time of the order of 2 sec.

At the beginning of each experiment the measuring system is calibrated at two or more temperatures in the liquid helium range, an average value of $\chi_c T_c$ being found (see equation (11)). After demagnetization, bridge settings are made at intervals of about 30 sec., so that the bridge reading immediately after cooling can be deduced by extrapolation. The value of T^* corresponding to the initial field and temperature is then known. From a series of such demagnetizations the relation between S and T^* is determined throughout the accessible range of temperature.

For the determination of Ξ at the low temperature reached on demagnetization, a solenoid is raised around the cryostat. Bridge settings are made alternately with and without a steady current flowing in the solenoid, the time of each reading being noted. The change in bridge setting produced by the steady field may be deduced by interpolation, thereby correcting for the steady rise of temperature caused by the 'heat leak', which amounted to about 40 ergs/min. Since the measuring system is very nearly astatic, except in so far as the balance is disturbed by the paramagnetism of the salt, the extra mutual inductance due to coupling with the solenoid is small. In order completely to eliminate this source of error, arrangements are made to close the solenoid circuit whenever a zero-field setting is being made, in such a way that the resistance in the solenoid circuit is nearly the same whether a current is flowing or not.

The magnitude of the steady field must be chosen so that the change in mutual inductance produced is not more than about 2 % of the total mutual inductance

due to the salt; otherwise the approximations made in § 2 will break down. It will be seen from the statement of the sensitivity given above that it is possible to measure a change of 2 % in a mutual inductance of $\sim 2000 \mu\text{H}$ with an accuracy of better than 1 %. The size of the appropriate steady field varies considerably with the salt used and with the temperature of the salt, for copper potassium sulphate the field required varies from about 50 G at 1°K to 1 G at 0.05°K . It is desirable that the amplitude of the measuring field shall be small in comparison with the steady field, although the effect of finite amplitude is easily seen to be, to a first approximation, zero. At very low temperatures it is necessary to allow for the vertical component of the earth's field, which may conveniently be determined by making measurements with the steady field alternately parallel and anti-parallel to it. The effect of the horizontal component may be neglected.

The heat leak is a surface effect and will tend to set up inhomogeneities of temperature within the salt, particularly at low temperatures where the thermal conductivity is very low (Garrett 1950*a*). It is therefore important that thermodynamic measurements should not be continued for too long a time after demagnetization (de Klerk, Steenland & Gorter 1949).

5. EXPERIMENTAL RESULTS FOR COPPER POTASSIUM SULPHATE

The spheroid of copper potassium sulphate was made from crystals supplied by B.H.D. Ltd., the purity not being stated. The mass of the spheroid was 12.6 g., the eccentricity 2.3:1 and the filling factor 0.94. Susceptibility measurements were corrected to spherical shape according to the formulae given in equations (14) to (17); on substituting for the known Curie constant (Reekie 1939) of the material, and for the demagnetizing factor and filling factor of the spheroid, equation (15) becomes

$$T_{\text{sphere}}^* = T_{\text{spheroid}}^* + 4.3 \times 10^{-3} \text{ }^\circ\text{K}. \quad (18)$$

The demagnetizations

The results of demagnetizations from various initial fields and temperatures are given in table 1. The entropy has been calculated from the statistical thermodynamical treatment given in the appendix, taking into account the anisotropic g factors of the cupric ion. No allowance has been made for the fact that the entropy in zero field at 1°K is already slightly below the value $R \ln 2$ holding for complete spin degeneracy. The demagnetizations are in two groups, those from a starting temperature of 2.0°K and those from 1.15°K ; the (S, T^*) curves for the two groups will be slightly displaced along the entropy axis, due to the extra lattice entropy at the higher temperature.

Measurements of Ξ

About 250 measurements of Ξ have been made at temperatures between $T^* = 0.018$ and 1°K , and a selection of the results is presented in table 2. The saturation corrections have been deduced by the method given in the appendix. The variation of Ξ with T^* is shown in figure 1. There is no region of temperature in which Ξ is at all constant, in contrast with the behaviour of potassium chrome alum (Garrett

TABLE 1

T_i	H (kG)	$R \ln 2 - S$ 10^5 erg units	T^*
1.187	1.015	1.64	0.590
1.203	1.35	2.81	0.386
1.179	1.35	2.93	0.359
1.175	1.35	2.95	0.347
1.139	1.35	3.13	0.324
1.122	1.35	3.29	0.313
1.109	1.35	3.30	0.348
1.125	1.705	5.15	0.291
1.155	1.73	5.02	0.257
1.120	2.03	7.25	0.179
1.158	2.25	8.36	0.155
1.118	2.37	9.89	0.136
1.152	2.505	10.40	0.134
1.136	2.71	12.48	0.112
1.117	3.045	16.21	0.0988
1.153	4.28	29.6	0.0612
1.122	6.77	73.5	0.0265
1.152	8.80	112.0	0.0177
1.136	10.82	161.0	0.0112
1.13	10.83	162.4	0.0103
1.975	2.03	2.36	0.976
1.977	2.71	4.38	0.354
1.996	2.71	4.19	0.365
1.987	2.78	4.38	0.367
2.002	3.05	5.17	0.353
1.992	3.38	6.38	0.241
2.001	4.06	9.19	0.181
2.002	4.40	10.60	0.150
2.002	4.74	12.30	0.136
2.002	5.08	14.0	0.129 ₆
2.002	5.42	15.9	0.108 ₃
2.005	6.09	20.0	0.089 ₆

TABLE 2

T^*	H (G)	\mathcal{M} (μH)	$\Delta\mathcal{M}$ (μH)	$\left(\frac{2\Delta\mathcal{M}}{3\mathcal{M}H^2}\right)^\odot$ $10^{-6} \times$	$\frac{2\psi}{\chi}$ $10^{-6} \times$	Ξ $10^{-6} \times$	χ
0.999	43.0	57.52	0.53	3.27	0.00	3.27	0.445
0.713	43.0	80.71	1.28	5.63	0.01	5.62	0.623
0.578	36.0	99.76	1.56	7.9	0.0	7.9	0.770
0.401	35.8	143.93	3.16	11.1	0.0	11.1	1.110
0.172	25.5	339.28	5.12	14.9	0.1	14.8	2.58
0.110	26.0	540.1	11.4	19.9	0.2	19.7	4.04
0.0789	19.0	766.3	11.4	26.3	0.3	26.0	5.64
0.0646	19.2	948.7	17.4	31.1	0.4	30.7	6.89
0.0319	11.5	2095.9	56.8	89.0	0.9	88.1	13.93
0.0234	3.39	3028.7	12.8	133.1	1.1	132.0	19.00
0.0189	2.37	3966.5	13.0	179	1	178	23.5

Note. Values of T^* , Ξ and χ are all reduced to spherical form (§3), as is the quantity shown in the fifth column. For the saturation correction $2\psi/\chi$, see appendix. \mathcal{M} is the mutual inductance, less that part of it not due to the salt; $\Delta\mathcal{M}$ is the change in the mutual inductance produced by a field H .

1948). Up to $0.5^\circ T^*$ the curve is tending towards a constant value, but between 0.5 and 1.0° the measured values decrease precipitately, due to the lattice specific heat, as has already been discussed (§3).

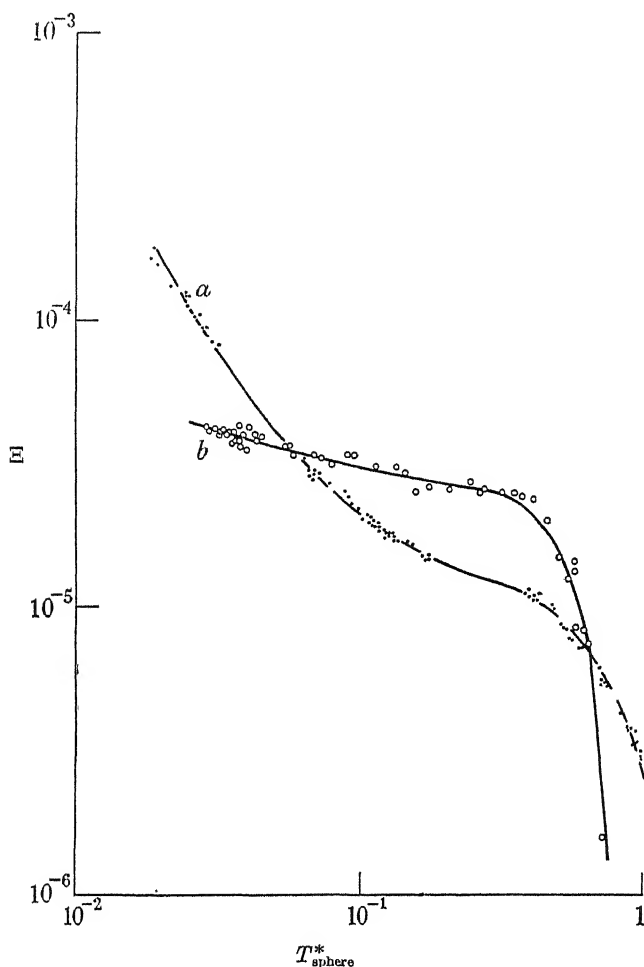


FIGURE 1.

Susceptibility and specific heat

Reekie (1939) has made absolute susceptibility measurements in the liquid helium range, finding that, within the limits of accuracy of his experiments, $\chi = 0.445/T^*$. From measurements of T^* below 1°K χ is therefore given by the formula

$$\chi = 0.445/T^*. \quad (19)$$

In figure 2 (curve *a*) Ξ^\dagger is plotted against values of χ derived in this way. It will be seen that, apart from the departure due to the specific heat of the lattice at high temperatures, the relation between the two quantities is linear down to $T^* = 0.025^\circ$. This suggests (equation (10)) that in this temperature range the susceptibility

follows a Curie-Weiss law, the specific heat having the form A/T^2 . From the graph it is found that

$$\frac{\Lambda}{A} = 8.87 \times 10^{-6} \text{ gauss}^{-2}, \quad (20)$$

$$\Delta = 0.034^\circ, \quad (21)$$

where $\Lambda = \chi(T - \Delta)$. Taking $\chi T = 0.445$ at 2.5° , $\Lambda = 0.438$ and

$$\begin{aligned} A &= 4.94 \times 10^4 \text{ erg deg./mole} \\ &= 6.0 \times 10^{-4} R. \end{aligned} \quad (22)$$

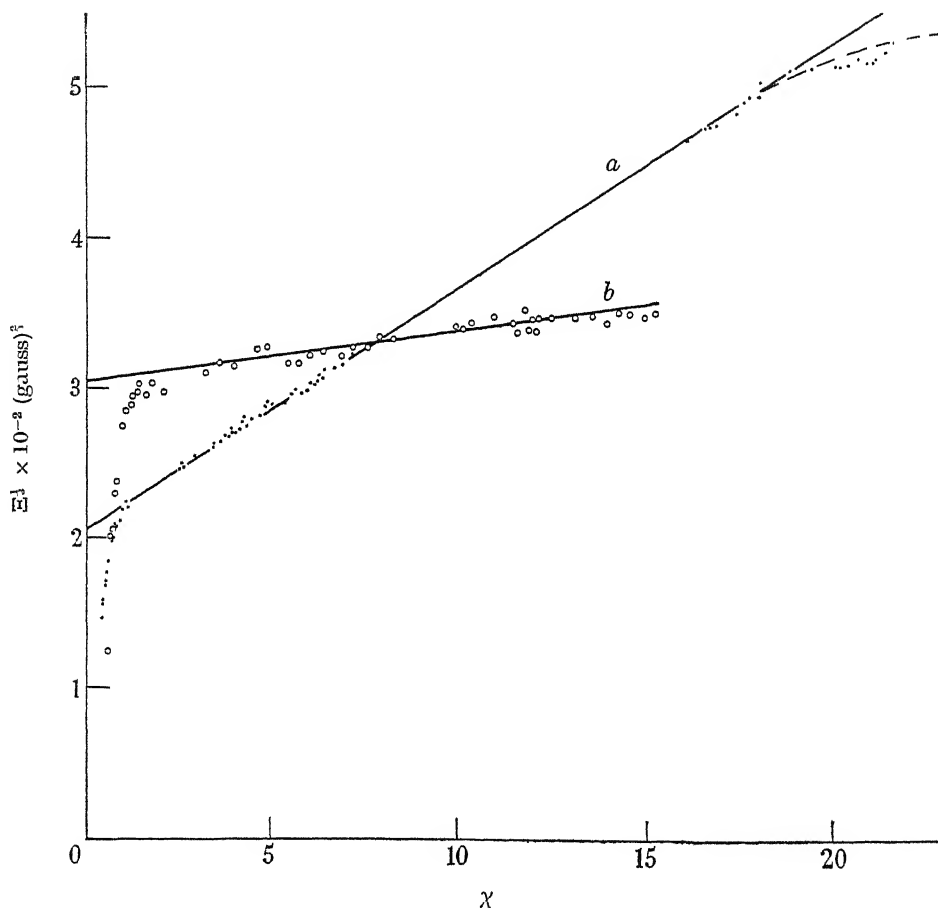


FIGURE 2.

So far there remains the possibility that χ and C are varying in some different way with temperature, the straight line in figure 2 being fortuitous. That this is not so is shown in figure 3 (curve a), in which, from the demagnetization data, S is plotted against $[T^*(2.5 - \Delta)/2.5 + \Delta]$, using the value of Δ found above (equation (21)). This again gives a straight line, as would be expected from equation (13). The gradient gives an independent estimate of A

$$A = 6.2_8 \times 10^{-4} R,$$

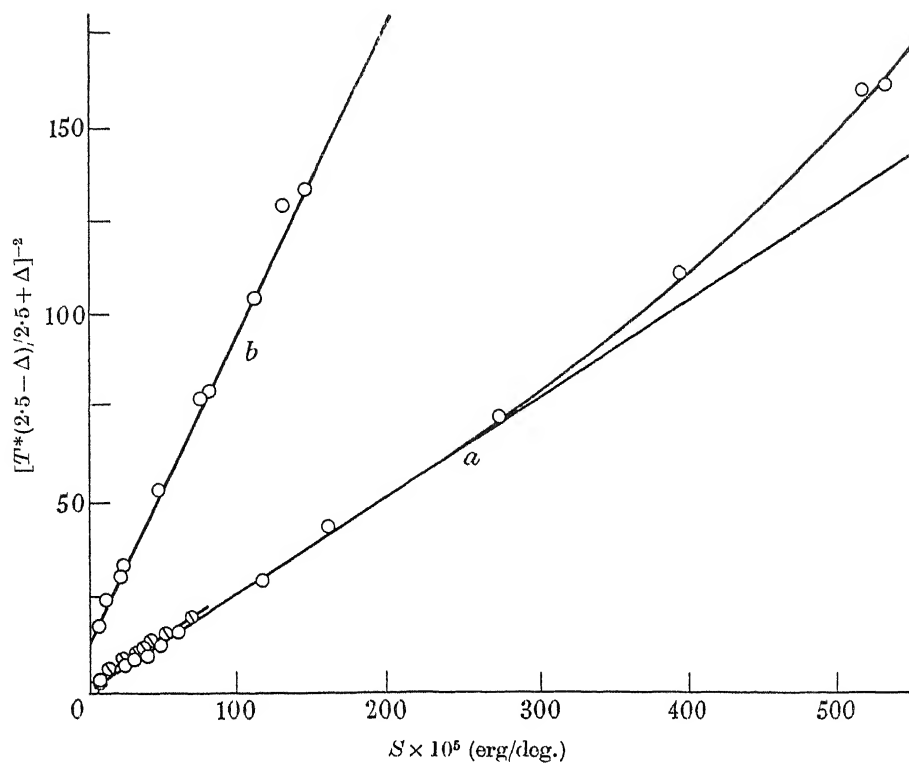


FIGURE 3. O, from 1 °K, ⊙, from 2.0 °K.

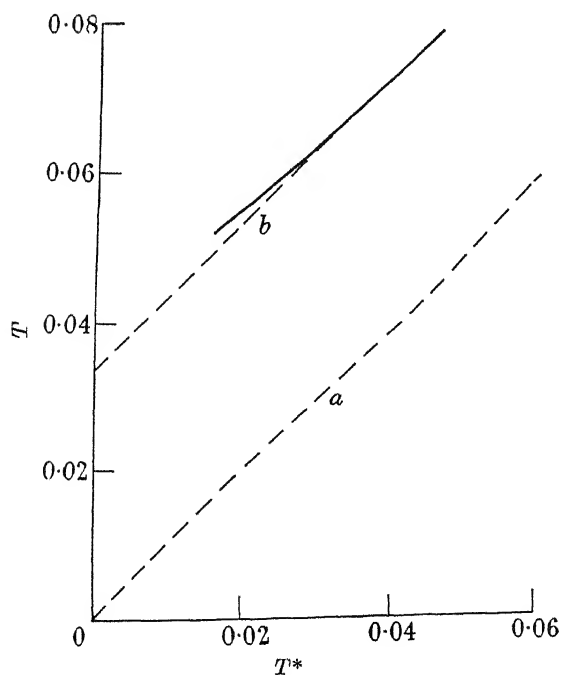


FIGURE 4. a, Curie law; b, Curie-Weiss law.

in satisfactory agreement with (22). It may be concluded therefore that, for temperatures down to $T^* = 0.025^\circ$,

$$T = 0.987T^* + 0.033^\circ \quad (23)$$

(compare equation (12)). For lower temperatures the general expression for T (equation (7)) must be used, differences in temperature being evaluated by graphical integration. The resulting T - T^* relation is shown in figure 4. In constructing this diagram it has been assumed that the Curie-Weiss law is still a good approximation at $T^* = 0.038$, so that there the true temperature is $T = 0.071^\circ$, the uncertainty in this fixed point is probably of the order of 2 millidegrees. It will be seen from figure 4 that the deviations from the Curie-Weiss law are in a ferromagnetic sense.

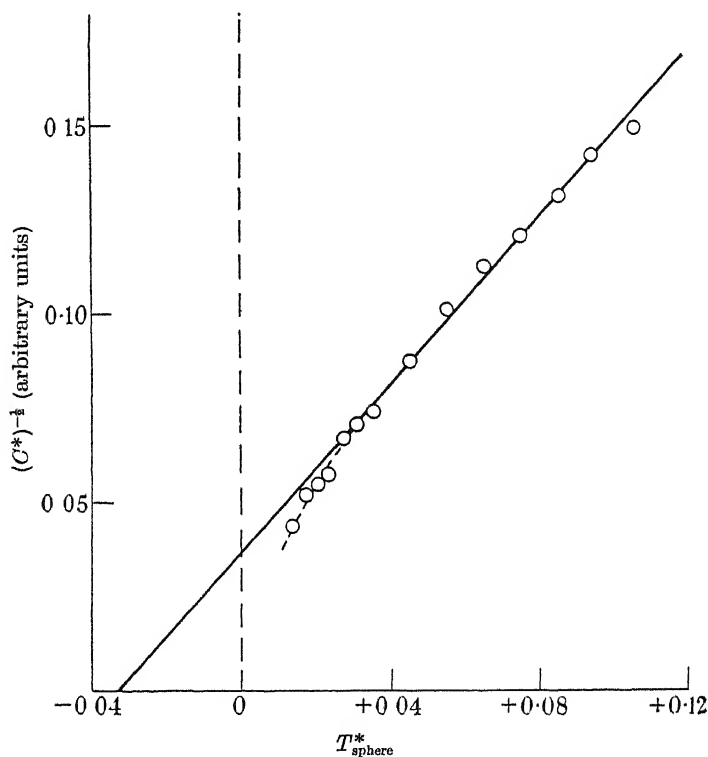


FIGURE 5.

Comparison with previous work

De Klerk (1946), by what is essentially a method of mixtures, has also found evidence for a Curie-Weiss law with this material, but his value of $\Delta(0.052^\circ)$ is in disagreement both with the demagnetization data presented here and with the measurements of Ξ . Further evidence comes from Ashmead's measurements[†] (1939) of the gamma-ray specific heat (dU/dT^*). If $C = A/T^2$ and $\chi = \Lambda/(T - \Delta)$, then

[†] I am indebted to Dr Ashmead for allowing me to make use of the more detailed account of his measurements contained in his dissertation.

$dU/dT^* \simeq A/(T^* + \Delta)^2$. In figure 5 experimental values of $(C^*)^{-\frac{1}{2}}$ are plotted against T . The intercept on the temperature axis is 0.033° , in satisfactory agreement with the value for Δ found in the present investigation (equation (21)).

6. EXPERIMENTAL RESULTS FOR THE DILUTED SALT

Mixed crystals were made from a solution of copper potassium sulphate and magnesium ammonium sulphate. The crystals were analyzed iodometrically and found to contain 0.0227 g. of copper per g., as compared with 0.144 g. in pure copper potassium sulphate. Thus in the diluted salt only 13.2 % of the lattice sites remained occupied by copper atoms, the other 86.8 % being replaced by diamagnetic magnesium.

The compressed spheroid made from these crystals was of mass 18.1 g. and eccentricity about 4:1. Values of T were corrected to spherical shape according to equation (15), substituting the appropriate quantities

$$T_{\text{sphere}}^* = T_{\text{spheroid}}^* + 0.90 \times 10^{-3} \text{ }^\circ\text{K.} \quad (24)$$

The demagnetizations

Table 3 shows the results of demagnetizations from various fields up to 10 kG. The entropy has been calculated exactly as for the undiluted copper potassium sulphate.

TABLE 3

T	H (kG)	$R \ln 2 - S$ (10^5 erg units)	T^*
1.203	3.38	17.2	0.12
1.201	4.06	24.6	0.0899
1.136	4.30	30.9	0.0644
1.194	4.74	34.0	0.0624
1.196	6.09	53.8	0.0416
1.142	7.20	79.9	0.0308
1.194	7.44	78.0	0.0322
1.192	8.73	104.4	0.0255
1.166	9.90	133.6	0.0217
1.192	9.95	129.0	0.0232

Measurements of Ξ

Ξ has been measured for the diluted salt at temperatures between $T = 0.028$ and 1°K . A selection of the results is shown in table 4, the saturation corrections being taken as before from the analysis given in the appendix. The specific heat per unit mass is about 25 times smaller at any temperature than for copper potassium sulphate; in consequence rates of warming are much faster, and thermodynamical measurements correspondingly more difficult and less accurate. The high-temperature decrease in Ξ due to the lattice sets in at a somewhat lower temperature than for the undiluted salt; this is because the lattice thermal capacity, expressed in terms of each gram-magnetic ion, is about eight times larger.

TABLE 4

T	H (G)	\mathcal{M} (μH)	$\Delta\mathcal{M}$ (μH)	$\left(\frac{2\Delta\mathcal{M}}{3\mathcal{M}H^2}\right)^\odot$ $10^{-6} \times$	$\frac{2\psi}{\chi}$ $10^{-6} \times$	Ξ $10^{-6} \times$	χ
0.687	75.3	9.96	0.37	4.4	0.0	4.4	0.65
0.439	51.9	15.75	1.32	20.6	0.0	20.6	1.01
0.277	46.4	24.90	2.16	26.6	0.0	20.6	1.61
0.177	46.4	39.13	3.38	26.4	0.1	26.3	2.51
0.106	37.6	65.10	4.45	31.4	0.3	31.1	4.18
0.0900	26.3	76.71	2.88	35.2	0.4	34.8	4.94
0.0726	32.7	95.73	5.42	34.1	0.7	33.4	6.13
0.0552	25.7	124.09	4.94	38.3	1.2	37.1	7.96
0.0368	20.1	189.63	5.38	43.6	2.7	40.9	12.08
0.0289	25.9	244.30	12.55	46.5	4.3	42.2	15.39

Susceptibility and specific heat

It will be assumed that Reekie's value for the Curie constant χT of copper potassium sulphate (0.445) holds also for the diluted salt. In figure 2 (curve b) Ξ^\ddagger is plotted against χ . The relation appears to be linear within the experimental accuracy, except for the decrease at high temperatures due to the lattice specific heat. Just as

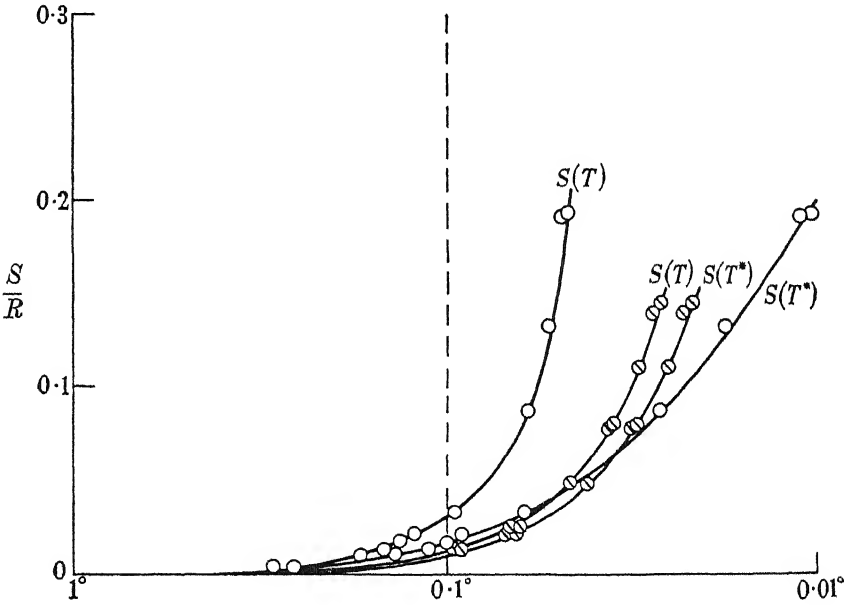


FIGURE 6. \circ , copper potassium sulphate; \odot , dilute salt.

for the undiluted salt this linearity suggests that the susceptibility is following a Curie-Weiss law (equation (10)). From the intercept and gradient it is found that

$$\begin{aligned}
 \Lambda/A &= 28.0 \times 10^{-6} \text{ gauss}^{-2} \\
 A &= 1.59 \times 10^4 \text{ erg deg./g. ion} \\
 &= 1.91 \times 10^{-4} R, \\
 \Delta &= 0.0048^\circ \text{ K.}
 \end{aligned}
 \tag{25}$$

(26)

The Curie-Weiss law is checked by plotting S against $(T^* + 0.0048)^{-2}$, as shown in figure 3 (curve *b*). This is again a straight line (equation (13)). From the gradient

$$A = 2.04 \times 10^{-4}R,$$

in reasonable agreement with (26). The average value

$$A = 1.98 \times 10^{-4}R \quad (27)$$

is probably correct to about 5 %. There are no detectable deviations either from the Curie-Weiss law or from the A/T^2 law down to the lowest temperature attainable ($T^* = 0.020^\circ$, $T = 0.025^\circ$). The dependence of the entropy upon T and upon T^* is shown in figure 6, which gives also, for purposes of comparison, the corresponding data for undiluted copper potassium sulphate.

Since this work was completed Benzie & Cooke (1949) have published results of measurements of the spin specific heat of various copper Tutton salts over a range of dilution, using a bridge method at liquid helium temperatures. A comparison with the results of the present investigation is deferred to part II (Garrett 1950*b*).

APPENDIX

In order to make use of the results of demagnetizations from fields up to 10,000 G at 1°K , it is necessary to be able to calculate the amount of entropy removed, taking into account the effect of paramagnetic saturation. For the measurement of Ξ it is also necessary to evaluate the saturation correction term ψH^3 in the expression for the magnetic moment (equations (5) and (6)). Once the energy level scheme in a field H is known, both problems may be solved by the standard methods of statistical thermodynamics. For free spins the Brillouin functions (Casimir 1940) may be used, but for crystalline paramagnetic salts the resulting expressions are likely to be considerably in error. In copper potassium sulphate the cupric ion behaves as a one-spin paramagnetic (see part II), but the Curie constant χT observed at liquid helium temperatures differs by 20 % from the one-spin value (Reekie 1939). The treatment given below takes into account the anisotropic g -factors of the cupric ion, due to Stark effect and spin-orbit coupling (Polder 1942; Pryce 1948); no allowance is made for magnetic or exchange interaction, or for the nuclear-electronic interaction (see part II), since there is reason to suppose that these effects are in this connexion negligible. The average magnetization for a single ion is first evaluated; this may then be used to determine the magnetization and entropy for a single crystal, and so also for a polycrystalline powder, by averaging over-all spatial direction of the applied field.

The Zeeman splitting, which would be simply $g\beta H$ for a free ion, varies in the crystal with the direction of the magnetic field. To take account of this g must be replaced by a tensor g_{ik} (Pryce 1948, 1950), the principal values of which may most precisely be determined by resonance experiments. For a one-spin paramagnetic there are two Zeeman levels, having energies $\pm \frac{1}{2}\beta \sqrt{\{(g_{ik}H_k)^2\}}$.† The partition function is

$$\begin{aligned} Z &= \exp[\beta \sqrt{\{(g_{ik}H_k)^2\}}/2kT] + \exp[-\beta \sqrt{\{(g_{ik}H_k)^2\}}/2kT] \\ &= 2 \cosh[\beta \sqrt{\{(g_{ik}H_k)^2\}}/2kT], \end{aligned} \quad (28)$$

† The summation convention for repetitive indices is assumed.

and the three components of magnetization are given by the equations

$$M_i = kT \frac{\partial}{\partial H_i} (\log Z), \quad (29)$$

from which it is found that

$$M_i = \frac{N\beta}{2} \frac{g_{mi}g_{mn}H_n}{\sqrt{\{(g_{ik}H_k)^2\}}} \tanh \left[\frac{\beta}{2kT} \sqrt{\{(g_{ik}H_k)^2\}} \right]. \quad (30)$$

This expression gives the average magnetization for one ion. For a polycrystalline sample it is now necessary to average over all directions of the applied field, it being of course assumed that the arrangement of crystals in the sample is perfectly random. For $H/T < 10,000$ G/deg. it is sufficient to expand the right-hand side of (30) as a power series in $\beta H/kT$, averaging the coefficients term by term. The entropy S is most conveniently derived by making use of the relation $(\partial M/\partial T)_H = (\partial S/\partial H)_T$. The resulting expressions are

$$M = \frac{N\beta\alpha}{4} \left[\overline{\sum_{xyz} l_x^2 g_x^2} - \frac{\alpha^2}{12} \overline{(\sum_{xyz} l_x^2 g_x^2)^2} + \dots \right], \quad (31)$$

$$S = -\frac{Nk\alpha^2}{8} \left[\overline{\sum_{xyz} l_x^2 g_x^2} - \frac{\alpha^2}{8} \overline{(\sum_{xyz} l_x^2 g_x^2)^2} + \frac{\alpha^4}{72} \overline{(\sum_{xyz} l_x^2 g_x^2)^3} - \frac{17\alpha^6}{11520} \overline{(\sum_{xyz} l_x^2 g_x^2)^4} + \dots \right], \quad (32)$$

as far as it is necessary to take them for the purposes of this paper. In equations (31) and (32) the l_x 's are direction cosines of H relative to the principal axes of g_{ik} , the g_x 's are the principal values of g_{ik} , and $\alpha = \beta H/kT$. The results of the averaging are

$$\begin{aligned} \overline{\sum l_x^2 g_x^2} &= \frac{1}{3} \Sigma g_x^2, \\ \overline{(\sum l_x^2 g_x^2)^2} &= \frac{1}{3} \Sigma g_x^2 + \frac{2}{15} \Sigma' g_x^2 g_y^2, \\ \overline{(\sum l_x^2 g_x^2)^3} &= \frac{1}{7} \Sigma g_x^6 + \frac{3}{35} \Sigma g_x^2 g_y^4 + \frac{2}{35} g_x^2 g_y^2 g_z^2, \\ \overline{(\sum l_x^2 g_x^2)^4} &= \frac{1}{9} \Sigma g_x^8 + \frac{12}{315} \Sigma' g_x^4 g_y^2 g_z^2 + \frac{18}{315} \Sigma' g_x^4 g_y^4 + \frac{20}{315} \Sigma g_x^2 g_y^6, \end{aligned} \quad (33)$$

the sign Σ' implying that each identical term is to be counted once only.

Microwave resonance experiments with copper potassium sulphate (Bleaney, Penrose & Plumpton 1949) indicate that the principal g values for the copper ion in this salt are

$$g_1 = 2.14, \quad g_2 = 2.04, \quad g_3 = 2.36,$$

substituting in equations (31) to (33) it is found that

$$\frac{\Delta S}{R} = -0.597\alpha^2[1 - 0.601\alpha^2 + 0.322\alpha^4 - 0.169\alpha^6 + \dots], \quad (34)$$

$$M = \frac{0.444H}{T} \left[1 - 1.81 \times 10^{-9} \left(\frac{H}{T} \right)^2 + \dots \right]. \quad (35)$$

It will be noticed that the coefficient of the first term in (35) agrees well with Reekie's experimental value (0.445).

Saturation corrections to Ξ

From equation (6) it will be seen that, in the parallel-field experiments, the proportionate change in bridge reading produced by a steady field H is $\frac{2}{3}(\Xi - 2\psi/\chi)H^2$.

From equation (35), $2\psi/\chi = 3.62 \times 10^{-9}/T^2$. Since this saturation term is temperature-dependent it is necessary to have some preliminary estimate of the true temperature before Ξ can be determined. Values of Ξ corrected by assuming $T = T^*$ are used, in conjunction with the measurements of $dS/d\chi$, to set up a relation between T and T^* . Some adjustment to the Ξ values may then be necessary, so that the procedure is in principle one of successive approximations. In practice, however, $2\psi/\chi$ is usually small in comparison with Ξ at all temperatures: for copper potassium sulphate at 1°K , $\Xi \sim 3 \times 10^{-6}$ and $2\psi/\chi \sim 4 \times 10^{-9}$; at 0.2°K , $\Xi \sim 1.5 \times 10^{-5}$ and $2\psi/\chi \sim 9 \times 10^{-8}$; at 0.06°K , $\Xi \sim 1 \times 10^{-4}$ and $2\psi/\chi \sim 1 \times 10^{-6}$.

I should like to express my gratitude to Dr J. Ashmead and Dr D. Shoenberg for much helpful advice and many stimulating discussions. My thanks are also due to Professor M. H. L. Pryce of Oxford University and to Professor D. R. Hartree on the theoretical aspects of the subject. I am indebted for help during the experiments to Dr E. B. Mendoza, with whom the magnetic cooling cryostat was first assembled, and to Mr B. B. Goodman, who also carried out the analysis of the mixed crystals.

For financial support I am indebted to the Department of Scientific and Industrial Research for a Maintenance Grant; to Trinity College, Cambridge, for a senior scholarship and research scholarship; and to the Worshipful Company of Skinners of London for the continuation of a Tonbridge School leaving exhibition.

REFERENCES

- Ashmead, J. 1939 *Nature*, **143**, 853.
 Benzie, R. J. & Cooke, A. H. 1949 *Nature*, **164**, 837.
 Bleaney, B., Penrose, R. P. & Plumptre, Betty, I. 1949 *Proc. Roy. Soc. A*, **198**, 406.
 Casimir, H. B. G. 1940 *Magnetism and very low temperatures*. Cambridge University Press.
 Casimir, H. B. G. & du Pré, F. K. 1938 *Physica*, **5**, 507.
 de Klerk, D. 1946 *Physica*, **12**, 513.
 de Klerk, D., Steenland, M. J. & Gorter, C. J. 1949 *Physica*, **15**, 649.
 Garrett, C. G. B. 1948 *Cérémonies Langevin-Perrin*. Paris.
 Garrett, C. G. B. 1949 *Nature*, **163**, 956.
 Garrett, C. G. B. 1950a *Phil. Mag.* **41**, 621.
 Garrett, C. G. B. 1950b *Proc. Roy. Soc. A*, **203**, 392.
 Gorter, C. J. 1948 *Physica*, **14**, 504.
 Polder, D. 1942 *Physica*, **9**, 709.
 Pryce, M. H. L. 1948 *Nature*, **162**, 538.
 Pryce, M. H. L. 1950 *Proc. Phys. Soc. A*, **63**, 25.
 Reekie, J. 1939 *Proc. Roy. Soc. A*, **173**, 367.

The nuclear specific heat in paramagnetic cupric salts at temperatures below 1°K

II. The nuclear specific heat

By C. G. B. GARRETT, PH.D

The Royal Society Mond Laboratory, University of Cambridge

(Communicated by Sir Lawrence Bragg, F.R.S.—Received 29 March 1950)

A theoretical discussion is given of the experimental results described in part I. Contributions to the low-temperature specific heat from dipole-dipole and exchange interaction are shown to depend linearly on the dilution factor when dilution has occurred in a purely random manner. From a comparison of the results for the diluted and undiluted salts it is shown that a part of the specific heat is independent of the degree of dilution. This part is identified with the nuclear specific heat expected from nuclear-electronic interaction within the atom. The interaction constants are evaluated and are of the order of 0.01°K . These results are in agreement with observations of the hyperfine structure in dilute copper salts made from paramagnetic resonance experiments. The effect of exchange interaction is also discussed.

1. INTRODUCTION

In part I (Garrett 1950) results have been presented of thermodynamic measurements made on copper potassium sulphate, and on a sample of the same salt diluted about seven times. The purpose of this part of the paper is to account for the observed specific heat constants (A) and Curie-Weiss constants (Δ) given in equations (21), (22), (26) and (27). The dilution factor for the second salt was 0.132 (§6).

2. ENERGY-LEVEL SCHEME

The level scheme of the cupric ion in the solid state has been discussed by Polder (1942) with particular reference to copper potassium sulphate. The ground state is a 2D state; there is one vacancy in the penultimate ($3d$) electron shell, and this may be treated as a single positive electron. The fivefold orbital degeneracy is completely removed by the electric crystalline field, and at and below helium temperatures only the lowest of the five orbital levels is appreciably populated. The introduction of electron spin makes each state doubly degenerate. As this spin degeneracy cannot be affected by the crystalline field (Kramers 1930), there can be no Stark specific heat anomaly in the region below 1°K . Susceptibility, anisotropy and resonance experiments are in general agreement with this picture.

Contributions to the low-temperature specific heat will come from dipole-dipole interaction and from exchange interaction. On the other hand, a spherical sample of a salt with only dipole-dipole interaction would have $\Delta = 0$ (van Vleck 1937), so that the observed non-zero values of Δ must be ascribed to exchange effects alone.

For copper potassium sulphate the dipole-dipole specific heat may be evaluated directly, and is found (de Klerk 1946) to be lower by a factor of 5 than the specific heat observed. The exchange specific heat cannot be calculated, because the exchange

integrals are unknown. Both dipole-dipole and exchange specific heats ought to tend to zero at infinite dilution. Since, however, dipole-dipole and exchange effects will still make some contribution to A at the dilution used, it is first necessary to discover how these contributions will depend on the degree of dilution, when that dilution has taken place in a purely random manner. It will be assumed (van Vleck 1937) that the various contributions to the specific heat are additive at temperatures large in comparison with the splittings.

3. DEPENDENCE OF INTERACTIONS ON DILUTION

Dipole-dipole interaction

The specific heat constant A_{DD} due to dipole-dipole interaction has been evaluated by van Vleck (1937), who concluded that

$$\frac{A_{DD}}{R} = \frac{2Q}{12} \left[\frac{\beta^2 g^2 \mathcal{N} S(S+1)}{k} \right]^2, \quad (1)$$

where β is the Bohr magneton, \mathcal{N} the number of ions in unit volume, g the Landé splitting factor and S the spin quantum number. Q represents the geometrical factor $\mathcal{N}^{-2} \sum_i r_{ij}^{-6}$, r_{ij} being the distance between the i th and the j th atom. In a regular dilution process Q will be unchanged and A_{DD} will be proportional to the square of the dilution factor ζ . But in the formation of mixed crystals there is always a fixed probability ζ that any particular site will be occupied by a paramagnetic ion, so that $Q \propto \zeta^{-1}$ and A_{DD} is only directly proportional to ζ .

Exchange interaction

The absolute magnitudes of A_{ex} and Δ are not accessible to calculation, but it is known that for 'isotropic' exchange

$$\frac{A_{\text{ex}}}{R\Delta^2} = \frac{3}{4f}, \quad (2)$$

where $2f$ is the number of neighbours between which exchange occurs. The exchange integrals are expected to depend on a high power of the interionic distance, so that in a regular dilution process the exchange specific heat should decrease exceedingly rapidly. But after random dilution there will still be, on the average, $2\zeta f$ exchanging neighbours, and hence, from equation (2),

$$\frac{A_{\text{ex}}}{R\Delta^2} \propto \zeta^{-1}. \quad (3)$$

If the exchange integral between two ions which happen still to be adjacent after dilution is unchanged, then Δ should be proportional to ζ (van Vleck 1937); from the experiments (equations I(21) and I(26)) $\Delta'/\Delta = 0.14$, compared with $\zeta = 0.13$. According to (3), A_{ex} should therefore be proportional to ζ in a process of random dilution. Both dipole-dipole and exchange interaction thus give rise to a specific heat depending linearly on the dilution.

These predictions can of course be verified by measuring the specific heat over a range of dilution. In the investigation described here only two salts have been studied, but more recently Benzie & Cooke (1949) have reported measurements at several dilutions, and their results confirm very satisfactorily the linear dependence of A upon ζ .

4. ANALYSIS OF THE SPECIFIC HEAT

It is now possible to discover, from a comparison of the specific heats of the diluted and undiluted salts, whether A is tending to zero or to a constant value at high dilutions. Experimentally it is found that $A'/A = 0.33$ (equations I (22) and I (27)), compared with $\zeta = 0.132$, so that the observed specific heat for the diluted salt is about 2.5 times as great as it would be if dipole-dipole and exchange interactions were alone responsible. It is therefore apparent that a part A_0 of the specific heat is independent of dilution, or depends on it but slightly. If then a represents that part of the specific heat of the undiluted salt which depends linearly on ζ ,

$$A_0 + a = 6.1 \times 10^{-4}R,$$

$$A_0 + 0.132a = 1.98 \times 10^{-4}R,$$

and hence

$$\left. \begin{aligned} A_0 &= 1.35 \times 10^{-4}R, \\ a &= 4.75 \times 10^{-4}R. \end{aligned} \right\} \quad (4)$$

The term A_0 is discussed in § 5. Part of a is due to dipole-dipole interaction and may be calculated by equation (1). Taking $S = \frac{1}{2}$ and $\bar{g}^2 = 4.77$ (Bleaney, Penrose & Plumpton 1949)

$$A_{DD} = 1.35 \times 10^{-4}R, \quad (5)$$

so that the exchange contribution is

$$\left. \begin{aligned} A_{\text{ex.}} &= 3.40 \times 10^{-4}R, \\ \Delta &= 0.034^\circ. \end{aligned} \right\} \quad (6)$$

These figures refer of course to the undiluted salt. It is interesting to compare (6) with equation (2), which holds for isotropic exchange; substituting the observed values,

$$\frac{3}{4f} = \frac{A_{\text{ex.}}}{R\Delta^2} = 0.29. \quad (7)$$

Crystallographic evidence suggests that the most probable number $2f$ of exchanging is 6, so that $A_{\text{ex.}}/R\Delta^2$ should be 0.25. However, if the exchange is in fact somewhat anisotropic, there is no difficulty at all in accounting for a value of $A_{\text{ex.}}/R\Delta^2$ somewhat in excess of $\frac{3}{4}$ (Opechowski 1948).

Benzie & Cooke, in the measurements referred to above (§ 3), have deduced the magnitude of the constant term A_0 by direct extrapolation of the results at several dilutions, using the ammonium, caesium and potassium copper sulphates. Their estimate of A_0 is $1.1 \times 10^{-4}R$, in reasonable agreement with (4).

5. INTERACTION WITH THE NUCLEUS

Gorter (1948) and Rose (1949) have drawn attention to the possibility of the existence of a nuclear hyperfine structure in the crystalline as well as in the gaseous state, and suggested that the effect might be used in experiments on the alinement of atomic nuclei. Garrett (1949), in connexion with results described in this paper, pointed out that this nuclear hyperfine structure should lead to a specific-heat anomaly of the well-known Schottky class, with a maximum at a temperature of the order of 0.01°K . Such a specific heat ought for copper potassium sulphate to form an appreciable fraction of the total specific heat at temperatures below 1°K , but the effect should be missing for most of the other materials used in magnetic cooling experiments. The unbalanced $3d$ electron shell may be imagined to set up at the nucleus of the copper atom a magnetic field of the order of 10^5 G ; the nuclear magnetic moment has $(2I + 1)$ allowed orientations relative to the electronic field, the energy separations being of the order of $\beta_N H_e/k$ ($= 10^{-2}$) degrees. As the material is warmed through this temperature range there will be a specific heat anomaly, on the high-temperature side of which the specific heat will be of the order of

$$(10^{-2})^2 R/T^2 \text{ per g.ion.}$$

Such a specific heat will be affected only very slightly by the process of magnetic dilution of the ions, in contrast with the behaviour of the dipole-dipole and exchange specific heats; and at infinite dilution the only specific heat remaining will be that due to the nuclear interaction. Thus the constant term (see §4) may be identified with the nuclear specific heat:

$$A_{\text{nuc.}} = 1.35 \times 10^{-4} R. \quad (8)$$

It is known that the hyperfine structure observed in the gaseous state (Ritschl 1932) is of the correct order of magnitude to account for a specific heat of this size, but in the solid state the problem is complicated by the effect of the electric crystalline field. The experimental value (8) is related to the interaction constants in the next section.

6. INTERPRETATION OF THE OBSERVED NUCLEAR SPECIFIC HEAT

The nuclear specific heat may be calculated in terms of one or more constants specifying the magnitude of the interaction between the electronic and nuclear systems. For a free ion the interaction is of the order of 10^{-2} cm.^{-1} and may be described by a cosine energy term $\Omega IJ \cos(IJ)$. In the crystal lattice, however, the electronic angular momentum L is decoupled from the electronic spin S , the total splitting of the orbital levels being $\sim 10^4\text{ cm.}^{-1}$. Introducing the nuclear interaction as a small perturbation, the first-order coupling between I and L is very small (of the order of $(10^{-2})^2/10^4\text{ cm.}^{-1}$), while the coupling between I and S should be to a first approximation unaffected by the crystalline field. But to take account of second-order perturbation effects due to coupling with the orbital angular momentum, the IS coupling must be regarded as anisotropic, i.e. no longer representable simply

by the cosine law. When the electric crystalline field has tetragonal symmetry the perturbation Hamiltonian may be written*

$$\begin{aligned}\mathcal{H} &= \Omega_1(I_x S_x + I_y S_y) + \Omega_2 I_z S_z \\ &= \frac{1}{2}\Omega_1[(I_x + iI_y)(S_x - iS_y) + (I_x - iI_y)(S_x + iS_y)] + \Omega_2 I_z S_z.\end{aligned}\quad (9)$$

In applying this expression to the cupric ion, which is regarded as a one-spin paramagnetic, $S_x S_y$ and S_z are the Pauli spin-matrix operators. For the copper nuclei ^{63}Cu and ^{65}Cu $I = \frac{3}{2}$, so that $I_x I_y$ and I_z are the analogous operators for a particle of spin $\frac{3}{2}$:

$$\left. \begin{aligned}I_x + iI_y &= \begin{pmatrix} 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & \sqrt{3} \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ I_x - iI_y &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ \sqrt{3} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & \sqrt{3} & 0 \end{pmatrix}, \\ I_z &= \begin{pmatrix} +\frac{3}{2} & 0 & 0 & 0 \\ 0 & +\frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{3}{2} \end{pmatrix}.\end{aligned}\right\} \quad (10)$$

The specific heat may now be evaluated by the diagonal-sum method (van Vleck 1937). The partition function is expanded as a power series in $1/T$, the terms of which involve expressions like $\text{spur}(\mathcal{H})$, $\text{spur}(\mathcal{H})^2$, The specific heat is then found from the relation $C = d/dT[T^2(d \ln Z/dT)]$. The term in $1/T$ vanishes, since it contains $\text{spur}(\mathcal{H})$, which is easily shown to be zero. The first non-vanishing term is that in $1/T^2$, and succeeding terms are not of interest since at high temperatures the $1/T^2$ term will predominate. It is found that

$$C = \frac{R \text{spur}(\mathcal{H})^2}{(2I+1)(2S+1)k^2 T^2}.\quad (11)$$

$\text{Spur}(\mathcal{H})^2$ may be computed in the unperturbed system of eigenfunctions, since the diagonal sum is invariant to a change in the representation. From (9) it may be shown after some reduction that

$$\begin{aligned}\text{spur}(\mathcal{H})^2 &= \frac{1}{4}\Omega_1^2\{\text{spur}[(I_x + iI_y)(I_x - iI_y)]\text{spur}[(S_x - iS_y)(S_x + iS_y)] \\ &\quad + \text{spur}[(I_x - iI_y)(I_x + iI_y)]\text{spur}[(S_x + iS_y)(S_x - iS_y)]\} + \Omega_2^2 \text{spur} I_z^2 \text{spur} S_z^2.\end{aligned}\quad (12)$$

* I am indebted to Professor M. H. L. Pryce for the above remarks on the influence of orbital angular momentum and for the Hamiltonian in equation (9). These qualitative considerations are justified by a second-order perturbation procedure discussed by him in a recent paper (Pryce 1950).

Substituting the expressions appropriate to copper (equations (10)), it is found from (11) and (12) that

$$C = \frac{5R}{16k^2T^2} (2\Omega_1^2 + \Omega_2^2). \quad (13)$$

This expression contains two interaction constants Ω_1 and Ω_2 , so that from the experimental result (equation (8)) it is possible only to deduce the value of $(2\Omega_1^2 + \Omega_2^2)$. It is known from microwave resonance experiments, however, that Ω_2 is very much larger than Ω_1 . If Ω_1 be neglected compared with Ω_2 , it is found that

$$\Omega_2 = 0.021^\circ\text{K} = 0.014\text{ cm.}^{-1} \quad (14)$$

Since this corresponds to a temperature of the order of the lowest reached in the present investigation, it is not surprising that the $1/T^2$ specific heat law was found to be quite a good approximation over most of the range of temperature studied.

It ought to be possible, just as in the case of the optical hyperfine structure, to interpret Ω_2 in terms of the electronic wave-functions and the nuclear magnetic moment, and hence to estimate the value of the magnetic moment from the experimental results. As yet, however, not enough is known about the quantum states of a paramagnetic ion under the influence of the crystalline field, and in any case the problem is of somewhat academic interest in view of the accuracy with which nuclear moments may now be measured by the resonance technique.

7. THE PENROSE EFFECT

Direct evidence for the hyperfine structure in the solid state has been provided by paramagnetic resonance experiments. Penrose (1949) found a hyperfine structure of four lines in the resonance spectrum of a dilute copper salt. Under the conditions of Penrose's experiment the salt is necessarily in a magnetic field so large that almost complete Paschen-Back quantization has taken place. Transitions between the two fourfold sets of levels can then occur, the selection rule being $\Delta m_I = 0$, and $(2I + 1)$ distinct resonances can be observed.

A detailed quantum-mechanical model can in principle be obtained by adding to the Hamiltonian (9) additional terms involving the magnetic field, but as yet there is no agreement between theory and experiment (Abragam & Pryce 1949; Broer 1949). It is, however, a simple matter to calculate from the line spacing a value for the low-temperature specific heat, without recourse to a microscopic theory. The partition function is set up for the four levels, the spacing of which is known from the resonance experiments, the specific heat is then derived, averaging over all field directions to take account of the observed variation of spacing with orientation. In virtue of the theorem of spectroscopic stability the result obtained must hold also for zero field, where the system of quantization is quite different. It is found (Pryce, private communication) that

$$A_{\text{nuc.}} = 1.1 \times 10^{-4}R, \quad (15)$$

in reasonable agreement with the value $1.35 \times 10^{-4}R$ deduced from the present investigation (8). In addition to the interaction with the nuclear magnetic moment, which alone is taken account of in (15), there will be a small additional effect due

to coupling with the nuclear electric quadrupole moment. This will appear in the resonance spectrum as a slight asymmetry in line spacing, which, however, it is difficult in practice to distinguish from the effect of incomplete Paschen-Back quantization.

8. CONCLUSION

Copper salts are practically unique among materials which have been used for magnetic cooling experiments, in that the nuclei of both the more abundant isotopes have non-zero magnetic moment. For most other salts the nuclei are of mass number $4n$ and so of zero moment; the nuclear degeneracy can be removed only in some other way than by magnetic interaction with the electron shells, and the nuclear specific heat maximum will be at a temperature much lower than any accessible by adiabatic demagnetization.

One exception to these remarks is the case of cobalt. Following Penrose's discovery of a hyperfine structure in the resonance spectrum of a copper salt, a similar but larger effect has been found in a dilute cobalt Tutton salt (Bleaney & Ingram 1949), from which it appears that the nuclear specific heat in cobalt salts should be rather more than ten times larger than in the case of copper. Measurements on the specific heat of a polycrystalline sample of cobalt ammonium sulphate below 1°K , using the method described in part I, have indicated a specific heat of the order expected from the resonance experiments. As cobalt ammonium sulphate is extremely anisotropic the measurements are being repeated with a single crystal.

In conclusion, it should be stated that the nuclear-electronic interaction can have no measurable effect on the susceptibility of the material, at any rate so long as the temperature is large in comparison with the splittings. For strong fields the electronic and nuclear systems will quantize separately (Paschen-Back state), the ratio of the susceptibilities being $(\beta_N/\beta)^2$, or about 10^{-6} ; and it follows from the theorem of spectroscopic stability (van Vleck 1932) that the same susceptibility expression which holds for Paschen-Back quantization must be correct at low field strengths, even though the system of quantization is then quite different.

I should like to express my gratitude to Dr J. Ashmead and Dr D. Shoenberg for much helpful advice and many stimulating discussions. My thanks are also due to Professor M. H. L. Pryce of Oxford University and to Professor D. R. Hartree on the theoretical aspects of the subject. I am indebted for help during the experiments to Dr E. B. Mendoza, with whom the magnetic cooling cryostat was first assembled, and to Mr B. B. Goodman, who also carried out the analysis of the mixed crystals.

For financial support I am indebted to the Department of Scientific and Industrial Research for a Maintenance Grant; to Trinity College, Cambridge, for a senior scholarship and research scholarship; and to the Worshipful Company of Skinners of London for the continuation of a Tonbridge School leaving exhibition.

REFERENCES

- Abragam, A. & Pryce, M. H. L. 1949 *Nature*, **163**, 992
 Benzie, R. J. & Cooke, A. H. 1949 *Nature*, **164**, 837.
 Bleaney, B., Penrose, R. P. & Plumptre, Betty, I. 1949 *Proc. Roy. Soc. A*, **198**, 406.
 Bleaney, B. & Ingram, D. J. E. 1949 *Nature*, **163**, 956.

- Broer, L. J. F. 1949 *Physica*, **15**, 673.
 de Klerk, D. 1946 *Physica*, **12**, 513.
 Garrett, C. G. B. 1949 *Nature*, **163**, 956.
 Garrett, C. G. B. 1950 *Proc. Roy. Soc. A*, **203**, 375.
 Gorter, C. G. 1948 *Physica*, **14**, 504.
 Kramers, H. A. 1930 *Proc. Acad. Sci. Amst.* **33**, 959.
 Opechowski, W. 1948 *Physica*, **14**, 237.
 Penrose, R. P. 1949 *Nature*, **163**, 992.
 Polder, D. 1942 *Physica*, **9**, 709.
 Pryce, M. H. L. 1950 *Proc. Phys. Soc. A*, **63**, 25.
 Ritschl, R. 1932 *Z. Phys.* **79**, 1.
 Rose, M. E. 1949 *Phys. Rev.* **75**, 213.
 van Vleck, J. H. 1932 *The theory of electric and magnetic susceptibilities*. Oxford. Clarendon Press.
 van Vleck, J. H. 1937 *J. Chem. Phys.* **5**, 320.

Emission of heavy fragments in nuclear explosions*

By D. H. PERKINS

1851 *Fellow, Imperial College of Science and Technology, London, S.W.* 7†

(Communicated by Sir George Thomson, F.R.S.—Received 4 April 1950.—

Revised 5 June 1950)

[Plates 7 to 10]

Observations have been made on the energetic nuclear fragments emitted in 'giant' cosmic-ray stars. The disintegrations were produced in the heavy nuclei of Ilford C2 and G5 emulsions exposed at high altitude. Methods are described for the identification of the fragments, both stable and unstable, with charges between three and ten units. It is found that fragment emission may be interpreted in terms of an asymmetric fission process, which is highly probable when the nuclear excitation energy becomes comparable with the binding energy. The average kinetic energies of the fragments are considerably greater than those to be expected from the Coulomb repulsion, a fact which may be explained in terms of specifically long-range nuclear forces. The fragments exhibit a peculiar tendency to be emitted in directions perpendicular to the incident star-producing particle. The mass of the emitted fragment does not depend appreciably on the size of star. At present there appears to be no comprehensive theoretical explanation of the precise mechanism of emission of such particles.

1. INTRODUCTION

Within the last few years a number of papers have appeared dealing with the distribution of charged particles emitted in the disintegration of nuclei excited to several hundred MeV energy. The excitation has been produced both by cosmic rays and by beams of artificially accelerated particles, and the phenomena have been studied in terms of the 'stars' in nuclear emulsions. These stars consist largely of tracks of protons and α -particles, and the distribution in energy and space of the low-energy

* The main substance of this paper was presented at the Como conference (September 1949) and at the Edinburgh conference (November 1949).

† Now at H. H. Wills Physical Laboratory, University of Bristol.

particles (below about 30 MeV) has been shown to be generally consistent with the evaporation model of the excited nucleus. According to such a model the total energy is statistically shared among the constituent nucleons before the emission of any one of them. The disintegrating nucleus is thus analogous to a drop of liquid; the velocity distributions of emitted protons, neutrons and α -particles are of Maxwellian type and determined by an effective nuclear temperature.

In the present paper we describe a type of phenomenon totally different from that of evaporation. It is characterized by the emission, in the more energetic nuclear explosions, of heavily charged fragments of very high kinetic energy and very low internal energy (Hodgson & Perkins 1949, Bonetti & Dilworth 1949, Sörensen 1949). The nature of such processes appears very complex, and the experimental data are not yet sufficient to explain their mechanism in detail. It was felt worth while, however, to present the available evidence in order to give an indication of some of the interesting new features of the phenomena.

A total of 78 heavy nuclear splinters, emitted from silver and bromine nuclei, have been observed in examination of 190 stars of more than 7 prongs, out of a total of 40,000 disintegrations in 100 μ Ilford C2 plates exposed at 11,000 ft.; and of 160 stars of more than 20 non-relativistic prongs, from a total of 8000 disintegrations in 400 μ Ilford G5 emulsions exposed in free balloons at an altitude of 80,000 ft. All plates were exposed vertically.

2. EXPERIMENTAL

2.1. *Identification of stable splinters*

The fact that occasionally particles heavier than α -particles may be emitted from stars has been known since the discovery of the well-known ${}^8\text{Li}$ 'hammer' track by Occhialini & Powell (1947).

The 'stable' splinters are first considered, that is, those which do not give rise to emission of heavy particles in the time required for them to come to rest in the emulsion. In discussing such splinters, it is necessary to distinguish between the long track of an energetic fragment sometimes emitted in the course of the disintegration, and the very short track produced by the low-energy recoil nucleus nearly always left over after the evaporation process has been completed. The 'recoil fragments' have nearly always a range of less than 10 μ (Harding 1949); in this paper observations are restricted to the more energetic fragments of ranges in excess of 30 μ .

A. *C2 emulsions*

The precise identification of a fragment in a star is in general difficult. It may be attempted by one or a combination of the following methods. Of the four methods, the first and second measure in effect secondary ionization (δ -rays) and the remaining two the primary ionization of the fragment (grain-density): (i) δ -ray frequency; (ii) δ -ray cut-off; (iii) gap counting; (iv) thin-down length. The first method is by far the most accurate, by reason of being a strictly comparative one

(i) *δ -ray frequency.* When fragments of high velocity have sufficient path-length in the emulsion, they may be detected and identified by the frequency of occurrence

of δ -rays along the tracks. δ -rays of 20 keV energy are clearly visible in C2 emulsions, corresponding to fragment velocities of $\beta > 0.14$ (velocity of light as unity). The δ -ray frequency, when plotted against the residual range of the particle, passes through a broad maximum, proportional to Z^2 , where Z is the charge on the particle. This gives a very simple and accurate method for determining Z . The maximum occurs at $\beta \sim 0.2$ for 20 keV δ -rays. Unfortunately, most fragments observed have

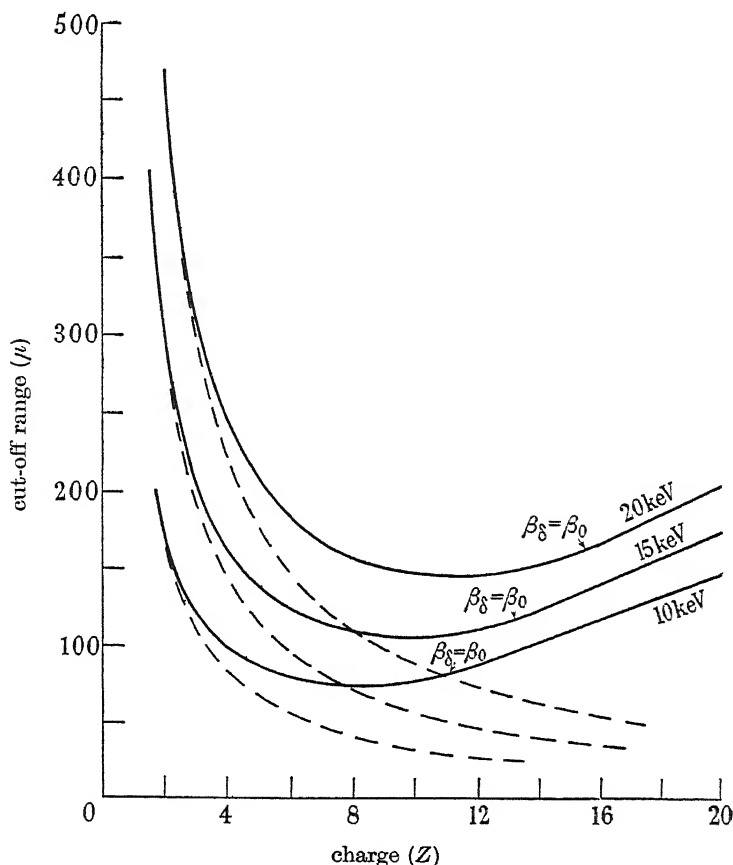


FIGURE 1. δ -ray cut-off range R_δ plotted against charge Z , for various δ -ray energies. ---- not allowing for thinning-down effect. — allowing for thinning down. For explanation of β_δ and β_0 , see text.

velocities well below this value, and the few high-velocity fragments do not usually have sufficient path-length in the emulsion for one to observe the maximum in the δ -ray frequency. In such cases, the observed frequency over a short section of the total range of the particle only enables one to put a lower limit to the charge.

(ii) δ -ray cut-off. The 'cut-off' in δ -rays of velocity greater than 2β should occur at a residual range, R_δ of the fragment given by the range-velocity relation $R_\delta = M/Z^2 f(\beta)$, where M is the mass of the fragment, if we neglect effects of reduction of effective charge due to capture of orbital electrons by a highly charged fragment as it slows down. If we assume $M = 2Z$, then for δ -rays of more than a given energy, $R_\delta \propto 1/Z$. This simple relationship is indicated by the dotted curves in figure 1. The

full-line curves in this figure were calculated making allowance for the effect of thinning down (see below), from which it is apparent that the range R_δ becomes very insensitive to changes in Z for values of $Z > 5$. Apart from this disadvantage, there remains the experimental uncertainty in determining R_δ . The latter must obviously be rather ill-defined because of (a) the impossibility of restricting the δ -ray counts exclusively to electrons of more than a certain definite energy, and (b) 'straggling' due to the fact that the orbital velocities of the electrons are not always small compared with β . Further, except in the case of very heavily charged fragments, the intensity of δ -rays along the track in the neighbourhood of the cut-off is so low that the average separation between consecutive δ -rays, and therefore the uncertainty in determination of the cut-off range, constitutes a considerable fraction of the cut-off range itself. The advantages of the method lie in the fact that an accurate δ -ray count taken over a great length of track is not required, and that it is applicable to all fragments of $\beta > 0.14$ of which the tracks lie completely in the emulsion. In many cases it serves as a useful check on charge values determined by other methods.

(iii) *Gap counting* In applying the two methods described below, both of which measure effectively the grain-density of the track of a fragment, it is to be noted that the shrinkage of the emulsions, during processing and drying, increases the apparent grain-density of steeply dipping tracks. This is owing to the fact that when a track is almost 'solid', a longitudinal compression of the silver filaments results in a lateral broadening. The thicknesses of tracks of α -particles from radiothorium 'stars' have been determined by planimeter measurements (see below). The increase in thickness with dip angle ϕ is found to be quite small for $\phi < 30^\circ$. Only the tracks of heavy fragments fulfilling this condition have been selected for measurement.

Since the tracks of α - and heavier particles are practically 'solid' for the velocities which we have to consider, discrimination between the two can best be performed by comparing the number of gaps in a given residual range, rather than the number of grains. As in grain-counting, the specification of a gap is almost entirely arbitrary, and depends on the optical system and observer. But if a certain convention is consistently applied, reliable results can be obtained. Figure 2 shows the results of gap-counting along the tracks of α -particles, 'hammer' fragments (^8Li) and tracks appearing heavier than α -particle tracks. There is quite good discrimination between α -tracks and those of heavier particles, for residual ranges exceeding 30μ . The gap-count plot shows clearly the existence of nuclei heavier than Li emitted from stars. All the tracks represented in this plot were selected according to the criteria that (a) $\phi < 30^\circ$, (b) they ended in the emulsion layer, (c) they had a gap-count plot differing significantly from that of an α -particle in the same star also fulfilling conditions (a) and (b). The actual determination of charge of heavy particles revealed in the gap-count plot is exceptionally difficult, but an attempt at this has been made as follows.

(iv) *Thin-down length*. A stripped, highly charged nucleus, in coming to rest in the emulsion, will capture orbital electrons as it slows down. This loss of charge results in a decrease in ionization, which, if sufficiently great, will lead to a reduction of thickness, or 'thinning-down' effect along the track. One can define a 'thin-down length' L as the average range at which the fragment, initial charge Z , captures its

first orbital electron (Freier, Lofgren, Ney & Oppenheimer 1948). Taking a simple Bohr model (Bohr 1941), it can be assumed that the fragment picks up an electron when it has velocity equal to that of the electron in its orbit. The fragment velocity is thus $\frac{1}{\nu} \left(\frac{z}{137} \right)$, when its instantaneous charge is z . Here ν is an effective total quantum number, as used in the Thomas-Fermi model. For the K -shell, $\nu \simeq 1$, as it is also for the outermost electrons ($z \simeq 1$). In the intermediate region $\nu \simeq Z^{\frac{1}{3}}$. Freier

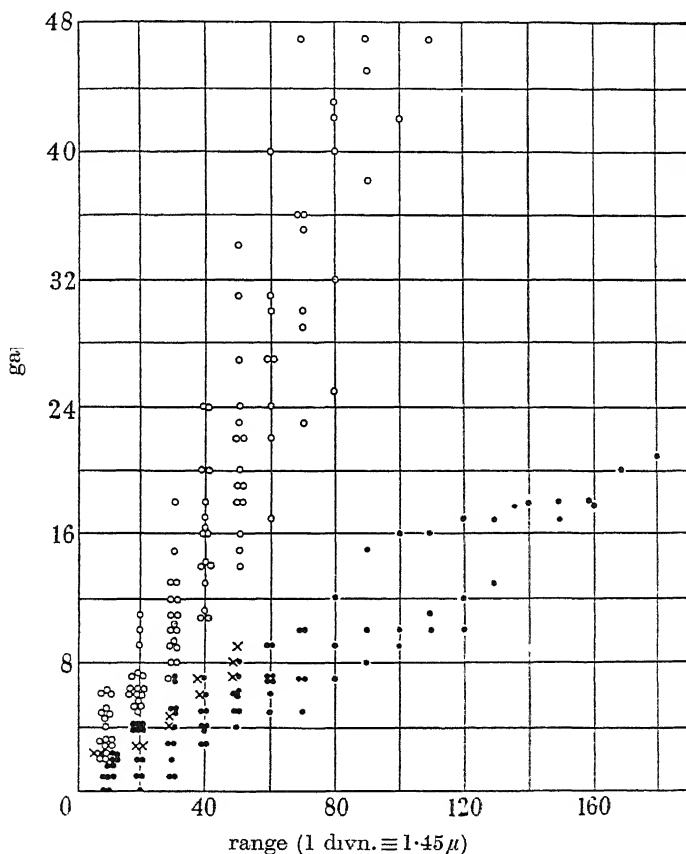


FIGURE 2. Results of gap-counting along tracks in C2 emulsion. O, α tracks; x, hammer tracks; ●, taper tracks.

et al. assumed $\nu = 1$ in all cases. This seems a fairly good approximation to use with a simple picture, provided Z is not too large. Assuming loss of charge to be a continuous process, and the mass $M = 2Z$, one obtains for the thin-down length, in microns,

$$L = 0.7Z^2, \quad (1)$$

by using the known range/velocity relation for protons and α -particles. This formula would not of course hold for high values of Z (e.g. uranium fission fragments) where the average value of ν is more nearly $Z^{\frac{1}{3}}$, and the index in (1) becomes only slightly greater than unity, and where, furthermore, nuclear stopping is very important. The relation (1) is probably not incorrect by more than a factor of two for $Z < 10$.

The experimental determination of L is rather difficult. First of all, the true thinning down may easily be confused with an apparent thinning down due to 'projected' δ -rays. The primary particle knocks-on δ -rays, with a maximum range decreasing with primary velocity. When the δ -rays are slow, they do not appear as distinct but merely as 'blobs' which are indistinguishable from the general 'buckled' structure of a solid track. They thus produce an overall thickening of the track, which decreases with decreasing residual range of the primary. This apparent thinning down can be clearly seen along the tracks of the heavy primary particles of the cosmic radiation. The projected δ -rays will make no contribution to the track thickness when their effective range (i.e. distance from point of origin to point of stopping) is smaller than this thickness, which is $\sim \frac{1}{2}\mu$ for α -particles towards the end of the range. We can therefore be quite certain that δ -rays have no thickening effect when their energies are less than 5 keV. The limit is more probably 10 keV, but in absence of a reliable relation between electron energy and projected range, it is difficult to be certain on this point.

The above considerations limit the region in which true thinning down could be observed, to L values less than about 70μ , and hence $Z < 10$.

Quite apart, however, from confusion arising because of slow δ -rays, one would not expect satisfactory measurement of L for high Z values for the following reason. The residual range of a particle, initial charge Z , instantaneous charge z , is from (1), $R = 0.7Zz$. If we assume, roughly, that the rate of change of track thickness, t , is proportional to the fractional rate of change of charge, then $dt/dR \propto 1/R$, so that most of the thinning would be at the end of the track.

In C2 plates, no tracks of very heavy fragments have been observed, which allow of an independent determination of charge, so that it is not possible to check formula (1). Planimeter measurements on five 'hammer' tracks due to ${}^8\text{Li}$ nuclei, gave a mean L value of $\sim 5\mu$, so that the formula appears correct in order of magnitude. In order to determine the thin-down lengths experimentally, drawings were made from enlargements of Kodak film negatives, which were themselves photomicrographs obtained with a projection microscope. The total magnification obtained was about $\times 9000$. The drawings were subdivided into short lengths and the area of each interval measured with a planimeter. In figure 3 is shown the result of planimetry of the track of a heavy fragment. The tapering effect is clearly demonstrated. L is estimated at roughly 25μ , giving $Z = 6$. The measurements in figure 4 were taken from the tracks of an α -particle, ${}^8\text{Li}$ hammer fragment and a heavier fragment, all occurring in the same star. Although, for the thickest track, there is a marked increase in thickness in the first 50μ from the end, projected δ -rays produce an additional thickening at longer ranges.

We summarize below the methods used for determining the charges of heavy fragments in various velocity ranges (all fragments must come to rest in the emulsion):

- (1) $\beta > 0.2$: δ -ray frequency.
- (2) $0.2 > \beta > 0.14$: δ -ray cut-off; thin-down length.
- (3) $\beta < 0.14$: if $R < L$, Z is indeterminate.

Since one requires a length of 'unthinned' track to compare with the thinned down portion, we have taken the limit at $R = 2L$. Further, all fragments of $R < 30\mu$

are excluded since one cannot discriminate between them and α -particles by a gap-count. We have therefore

- (i) $R > 2L$ and $> 30\mu$, Z determined by thin-down length.
- (ii) $\left. \begin{matrix} R < 30\mu \\ R < 2L \end{matrix} \right\} Z$ indeterminate.
- (iii) L appearing greater than 70μ ; $Z \geq 10$ but otherwise indeterminate.

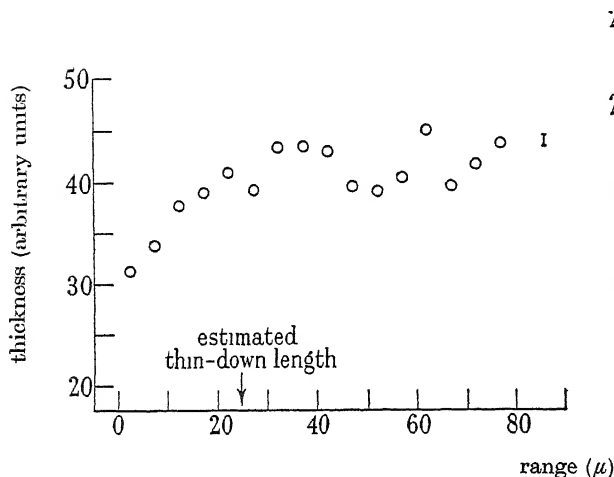


FIGURE 3

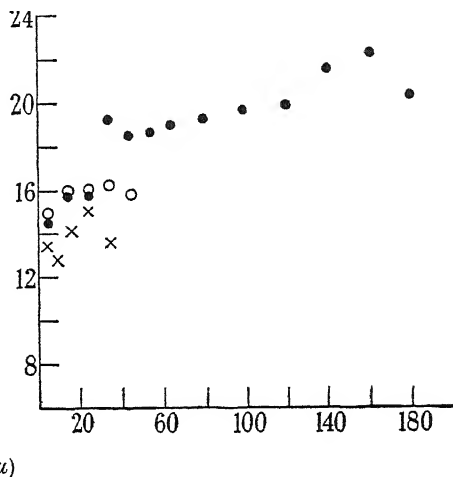


FIGURE 4

FIGURE 3. Planimeter measurements along highly magnified image of track of fragment of charge $Z \simeq 6$. I, error on planimeter reading.

FIGURE 4. Comparison of planimeter measurements on tracks of α -particle, ${}^8\text{Li}$ hammer, and heavier particle emitted from a single star. \times , α -particle; \circ , hammer (${}^8\text{Li}$); \bullet , heavy particle.

The following example is illustrative of the methods above:

13-branch star (see figure 11, plate 10).

Track (1). α -particle (range = 95μ): gaps in 6.3μ intervals from end:

2, 2, 4, 2, 4, 3, 2, 1, 3, 2, 2, 3, 2, 4.

Track (2). Range = 250μ .

- (a) Gap count: 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 2, 1, etc. Therefore heavier than α -particle.
- (b) δ -rays of 20 keV observed (hence $\beta > 0.14$) up to 160μ from end of track. Thus $Z \geq 8$.
- (c) δ -rays in first 63μ of track = 18. Maximum in δ -ray frequency along track of fast α -particle is $\sim 2.0 \delta$'s/ 63μ interval. Therefore $Z \geq 6$.
- (d) Thin-down length = 50μ ($R > 2L$), hence $Z \simeq 9$. We conclude that $Z = 9$, the fragment having energy 220 MeV and velocity $\beta = 0.17$.

Track (3). Range = 46μ ($R > 30\mu$).

- (a) Gap count: 2, 1, 0, 0, 0, 0, 0. Therefore heavier than α -particle.
- (b) No δ -rays, hence $\beta < 0.14$.
- (c) Thin-down length = 16μ ($R > 2L$), giving $Z = 5$, energy 28 MeV and $\beta = 0.08$.

When the charge Z and residual range R of a fragment are known, the energy can be calculated from the following formula, which takes account of thinning down:

$$\frac{E}{Z} = 0.20(ZR)^{0.67} - 0.07 Z^2, \quad (2)$$

where E is in MeV, R in microns. The full curves of figure 1 were calculated from this formula. If β_δ is the primary velocity for δ -ray cut-off, the energy E_δ and hence cut-off range R_δ can be found for any given value of Z . Formula (2) holds as long as β_δ exceeds the pick-up velocity $\beta_0 = Z/137$. When $\beta_\delta < \beta_0$, the instantaneous charge z of the fragment at the cut-off is given by $\beta_\delta = (z/137)$. Thus $R_\delta = Z/z \times (\text{thin-down length of fragment of initial charge } z, \text{ mass } 2z)$. Since z is fixed for a given β_δ , R_δ increases linearly with Z in the region $\beta_0 > \beta_\delta$.

B. G5 emulsions

The electron-sensitive G5 emulsions are, generally speaking, less satisfactory than the less sensitive C2 emulsions for recording of star fragments. G5 emulsions were employed chiefly with a view to investigating possible connexions between splinter production and penetrating showers (§7).

The great advantage of electron-sensitive plates is that δ -ray counting along tracks of high-velocity splinters is much easier than in C2 plates. This is offset by the much poorer discrimination between heavily ionizing fragments in the more sensitive emulsions. Since all tracks of such particles are completely 'solid', the gap-count method for differentiating between α -particles and splinters cannot be used. For the same reason, the thin-down effect along tracks of all but the very highly charged fragments is almost obscured. Thus a large proportion of the low velocity fragments in G5 plates escape detection.

2.2. Identification of unstable splinters

In the preceding section has been considered the problem of identifying heavy fragments from stars, when their internal energy is low so that they remain stable against heavy-particle emission during the time required for them to come to rest in the emulsion. Since this time is of the order of 10^{-12} sec., it would appear that the only activity in general possible for these fragments after coming to rest is either γ -emission involving high spin change (octopole or higher order radiation) or β -decay (followed possibly by delayed heavy particle emission, as in uranium fission). None of these processes is in general observable in C2 emulsions; β -decay is inferred from the subsequent α -disintegration in the case of ^8Li 'hammer' tracks. In G5 emulsions, such β -decay is directly observable. One example has been found of a very heavy fragment ($Z \approx 8$) which emits a β -particle after coming to rest.

However, we may expect that sometimes a fragment may be emitted with sufficient internal energy to give rise to heavy-particle emission directly. The life-time for such processes will almost certainly be less than 10^{-14} sec., so that the fragment will not leave an observable track in the emulsion before it disintegrates. In the few cases observed where the track of a heavy fragment ends in a star of heavy particles, it is not at present possible to say whether the nucleus itself undergoes

disintegration after a remarkably long lifetime, or whether it produces a star by impact with one of the light emulsion nuclei (Lovera 1947; Schopper 1947, Hodgson & Perkins 1949).

If the excitation energy of the fragment is just sufficient to overcome the binding energy of its constituent particles, the latter will appear to emanate from the star in a narrow cone. The fragment may therefore be identified in those instances where an energy level of the excited nucleus happens to lie just above the disintegration energy, and where a disintegration solely into charged particles is a highly probable process. A rough calculation shows that, owing to the relatively wide spacing of energy levels in light nuclei in such excited states, the probability of observing these cones will generally be quite small.

A complete statistical analysis of the tracks from stars, for cones of all sizes, has not yet been attempted; only the simpler analyses for two- and three-particle cones have at present been carried out.

(i) ${}^8\text{Be} \rightarrow 2 {}^4\text{He}$ disintegration

A well-known example of a nucleus fulfilling the above conditions is ${}^8\text{Be}$, in which the binding energy of the two α -particles in the ground state is small and negative. When ${}^8\text{Be}$ is emitted from a star, it will possess either even or odd angular momentum, l . If l is even, the nucleus will disintegrate spontaneously into two α -particles, presumably after surmounting the potential barrier of the parent nucleus. The lifetime of ${}^8\text{Be}$ has been estimated by Wheeler (1941) at 10^{-16} sec. If l is odd, the two- α disintegration is forbidden and only γ -transitions are possible. When l is small, the lifetime for γ -emission is of the order of 10^{-13} sec., so that the ${}^8\text{Be}$ will disintegrate into two alphas before it has had time to leave an observable track in the emulsion. Only in the case where l is large and odd, and γ -emission involves a high spin change, will the nucleus come to rest before undergoing de-excitation; this will be followed immediately by the α disintegration, giving rise to a 'hammer' track. In electron-sensitive plates, these will appear as hammer tracks unaccompanied by a β -particle, and hence not due to ${}^8\text{Li}$.

Where the ${}^8\text{Be}$ disintegrates immediately after leaving the nucleus, the resulting α pairs should produce an anomalous angular and energy distribution of α -particles from stars.

The observed number of pairs of α -particles, each lying completely in the emulsion, with angular separations in azimuth $\Delta\theta$ and dip $\Delta\phi$, have been compared with the numbers to be expected from a random distribution. Table 1 refers to a sample of stars in C2 emulsions.

TABLE 1

star prong number	17-31	12-16	7-11	total
$\Delta\theta, \Delta\phi < 10^\circ$ { observed	7	6	1	14
{ expected	2.1	1.5	0.4	4
$\Delta\theta, \Delta\phi < 20^\circ$ { observed	9	13	3	25
{ expected	8.5	6.4	1.5	16.4

The expected numbers of random coincidences were computed using the observed dip angular distribution of all α -particles coming to rest in the emulsion. We see that,

for $\Delta\theta, \Delta\phi < 10^\circ$, the observed number of pairs is significantly greater than the expected number. In figure 5 is shown the frequency distribution of the ratio x of the energies of each pair of α -particles separated by an angle $\alpha < 10^\circ$. It is quite different from the random distributions to be expected if the energies were independent (a Maxwellian type of energy spectrum was assumed for evaporating α -particles; Perkins 1950). We assume that *all* the α pairs with separation $\alpha < 10^\circ$ arise from disintegration of ^8Be , and calculate the excitation energy from the formula

$$\epsilon = \frac{1}{2}(E_1 + E_2 - 2\sqrt{E_1 E_2} \cos \alpha),$$

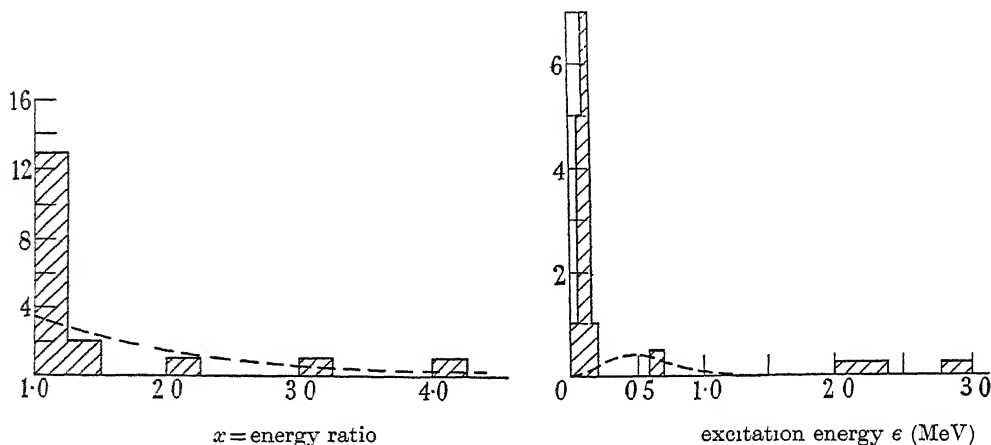


FIGURE 5. The histogram indicates the observed frequency distribution of the ratio x of energies of pairs of α -particles of separation $< 10^\circ$. ---- 'random' distribution $w(x) dx = 54x/(1+x)^4 dx$.

FIGURE 6. Distributed in excitation energy, ϵ , of ^8Be fragments emitted from stars. The dotted curve is the expected apparent distribution due to unassociated 'random' α pairs.

where E_1, E_2 are the energies of the α -particles. The distribution of ϵ is shown in figure 6. The ground-state of ^8Be is known to be unstable to α -disintegration by $\sim 100 \text{ keV}$, and indeed most of the values of ϵ are grouped around this figure ($100 \pm 30 \text{ keV}$). The observed spread can be attributed to straggling and scattering along the α tracks, errors in measurement of α , and to a small extent, random coincidences. The fractional error in ϵ due to straggling is of the same order as the straggling itself, i.e. $\sim 3\%$ for the energies considered; that due to uncertainties in α is up to twice the fractional error in α . The latter is estimated at 10% on the average, so that we expect a total 'straggling' in ϵ of at least $\pm 25\%$. This is reasonably consistent with the observations.

The first excited state (also of $l = 0$) of ^8Be is known to be $2.8 \pm 0.8 \text{ MeV}$; three of the four remaining values of ϵ are attributable to this state. The possibility cannot, however, be excluded that such a small number of events may be due to random coincidences. The apparent distribution in ϵ due to these is shown by the dotted curve. It seems likely therefore that the three values of ϵ between 2 and 3 MeV are genuine examples of excited ^8Be .

The angular distribution of the α -particles in $l = 0$ states should be isotropic in the centre-of-mass system, and for values of $\epsilon < 150$ keV (ground state) this is in fact found to be true. On the other hand, the condition $\alpha < 10^\circ$, whilst not important for disintegrations from the ground state, restricts observations in the case of the 3 MeV state to those pairs in which the α 's fly off in nearly the same direction as that of the original ${}^8\text{Be}$ nucleus. If we apply a rough correction factor, we then obtain the following data for the emission frequency:

	number
ground state (100 keV)	14 ± 3
excited state (3 MeV)	20 ± 11
total	34 ± 11

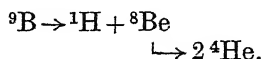
Of course we have no information about possible higher excited states; the angle of divergence of the α pair would be so great as to be always obscured by random coincidences.

(ii) ${}^9\text{B} \rightarrow 2 {}^4\text{He} + \text{H}^1$ *disintegration*

Up to the present, three examples have been observed of the concentration in a narrow cone of two α -particles and a proton, all ending in the emulsion. The expected number of random coincidences of this type is less than 0.1. The distinguishing feature of this process is that, as can be seen in figure 10, plate 9, the α 's and proton have nearly the same range, and hence practically equal velocities. The following data are from one example:

	energy (MeV)	ϕ° (dip angle)	θ° (azimuth)	velocity ($\times 10^9$ cm./sec.)
alpha	18.2	+3	2	2.87
alpha	19.0	-3	4	2.93
proton	3.95	-1	0	2.67

It seems highly probable that these cones represent the spontaneous disintegration of ${}^9\text{B}$, assumed to be of the form



Denoting by η the instability of the ${}^9\text{B}$ in its ground state with respect to disintegration into a proton and ${}^8\text{Be}$, ψ the angle between the direction of the ${}^8\text{Be}$ and the proton in the plate system, E_{Be} and E_{H} their energies, we obtain

$$\eta = \frac{4}{3} [2E_{\text{H}} + E_{\text{Be}}/4 - 2\sqrt{E_{\text{H}}E_{\text{Be}}}/2 \cos \psi].$$

We thus find $\eta = 50, 90$ and 110 keV in the three cases. These values agree well with that from mass tables of 160 ± 90 keV (Rosenfeld 1948).

(iii) ${}^{12}\text{C} \rightarrow 3 {}^4\text{He}$ *disintegration*

Two probable examples of the emission of excited carbon 12 as a group of three α -particles have been observed. The expected number of such cones, that is, random coincidences of an α with a two- α pair, is 0.5. One cannot of course draw definite conclusions from such a small number of events. They are phenomenologically not striking as the cases of ${}^9\text{B}$ 'cones' described above, the angular divergence being

much greater ($\sim 10^\circ$ semi-angle). From the binding energy of α -particles in ^{12}C , and their energies and directions in the laboratory system, we calculate the excited state of ^{12}C and obtain values of 8.7 and 9.3 MeV in the two cases. The error is ~ 0.5 MeV and arises chiefly from errors in measurement of angles of dip. The known levels in ^{12}C , above the disintegration energy (7.4 MeV), are 9.5, 16.1 MeV, etc. The disintegration in each case is found to proceed by the stage $^{12}\text{C} \rightarrow {}^8\text{Be} + {}^4\text{He}$, where the ${}^8\text{Be}$ is in the ground state.

(iv) $^{16}\text{O} \rightarrow 4 {}^4\text{He}$ disintegration

One example has been observed of the emission from a star of two ${}^8\text{Be}$ α pairs. All four α tracks had practically zero dip-angle. One α pair gave an ϵ value of 90 keV (ground state), the other 2.25 MeV (first excited state). The separation between the ${}^8\text{Be}$ fragments was 28° , their kinetic energies 44.9 and 49.8 MeV. From these data, one can calculate the excitation energy of an ^{16}O fragment breaking up into two ${}^8\text{Be}$ nuclei. The level found is 22.85 MeV; owing to the very favourable geometrical conditions for measuring the α tracks, the error in this figure is only a fraction of a MeV. It agrees well with the known level in ^{16}O at $23.9 \begin{Bmatrix} +0.5 \\ -1.5 \end{Bmatrix}$ MeV, determined by Newson studying the yield from the $^{14}\text{N}(d, n)^{15}\text{O}$ reaction.

3. ANALYSIS OF RESULTS

3.1. Frequency distribution of fragments from stars

Owing to the impossibility of identifying short-range fragments of low charge in G5 emulsions, we restrict our observations first of all to the results from C2 plates. Numbers of fragments, both stable and unstable, are given in table 2, where the stars have been divided into different ranges of prong number. The estimated true frequencies of fragments are given in brackets; they are obtained by applying correction factors for (a) geometrical loss of long tracks out of the emulsion layer, (b) restriction in dip angle ($\phi < 30^\circ$), on the assumption that the distribution of tracks is isotropic. This point will be further discussed below. The loss factors for particles of a particular charge were obtained from the mean observed range of particles of that charge.

TABLE 2

prong no.	7-11	12-16	17-21	22-30	total
no. of stars	58	75	46	10	189
$Z=3, 4$	2 (7 \pm 5)	10 (38 \pm 13)	11 (40 \pm 12)	2 (7 \pm 5)	25
$Z=5, 6$	2 (2.5 \pm 2.5)	5 (13 \pm 5.5)	5 (13 \pm 5.5)	1 (2.5 \pm 2.5)	12
$Z=7, 8, 9, 10$	0 (0)	1 (4 \pm 4)	3 (12 \pm 7)	1 (4 \pm 4)	5
δ -ray tracks	0	0	3	0	3
unidentified	2	2	5	0	9
total frequency	5/58	18/75	27/46	4/10	—
(observed)					
corrected frequency	0.20 \pm 0.10	0.75 \pm 0.20	1.5 \pm 0.4	1.4 \pm 0.7	—
per star					
mean charge on ejected fragments	4.1	4.4	4.7	5.3	—

The row marked ' δ -ray tracks' refers to those fragments of high velocity passing out of the emulsion layer, on which the δ -ray count indicated a charge significantly greater than 2. The 'unidentified' tracks were those along which thinning down was unobservable or very uncertain. The number of fragments per star is seen to increase very rapidly with prong number. Although the statistical error is large, the average fragment charge does not appear to change rapidly with excitation energy.

The number of fragments of various charge, independent of size of star, are given in table 3.

TABLE 3

Z	3	4	5	6	7	8	9	10
observed	7	18	7	5	1	2	1	1
corrected	14 ± 5	78 ± 19	17 ± 6	14 ± 6	4 ± 4	8 ± 5.5	4 ± 4	4 ± 4

Here we see immediately the great preponderance of fragments of charge 4 (mostly ^8Be) as compared with charge 3. This applies even if we halve the corrected number of ^8Be fragments, in order to consider only those cases of ^8Be in the ground state. This seems to support the view that not only α -particles but also ' α -particle nuclei' may exist temporarily as subunits inside the nucleus (Wheeler 1941). In this table, all hammer tracks are assumed to be due to ^8Li fragments.

Comparison has been made of fragment frequencies in stars in C 2 and G 5 emulsions. Allowing for the different sensitivities and thicknesses of the two types, the frequencies in stars of comparable sizes are found to be in good agreement.

Table 4 gives the frequency of emission of fragments as a function of excitation energy. We must differentiate here between the *total* energy released in the explosion, i.e. the kinetic and binding energies associated with all the ejected particles, except the shower particles, and the energy which goes into surface and volume vibrations of the nucleus. It is reasonable to assume that, like the shower particles (§ 3.4) fast protons and neutrons knocked directly out of the nucleus by the incident particle have no part in the 'fission' process. The energy, U , appropriate to the latter has therefore been estimated by considering only the so-called 'evaporation' particles, i.e. those with energies below 25 to 30 MeV.

TABLE 4

prong no. in $\begin{cases} \text{C 2} \\ \text{G 5} \end{cases}$	7-11	12-16	17-21	22-30
	8-13	14-19	20-25	26-35
charge carried by emitted particles (except shower particles)	14	24	31	40
total excitation energy (MeV)	900	1700	2700	4000
U = energy required to release particles of energy below 25 MeV	370	600	950	
			1.5 ± 0.4	1.4 ± 0.7
corrected (stable) fragment frequency per star	0.2 ± 0.1		0.75 ± 0.20	
			1.5 ± 0.3	
approximate initial temperature (in MeV)	$5\frac{1}{2}$	7	9	

The binding energy of a silver or bromine nucleus is ~ 750 MeV. When U is less than or of the order of W , the fragment frequency increases very rapidly with U . For $U \sim 900$ MeV, fission has become the dominant process and there appears to be no

further significant increase in frequency for higher excitation energy. We note that the disintegrations involving more than 20 prongs in C2 emulsions lead to complete breaking up of a bromine nucleus, and almost complete disintegration of silver. Thus, when the prong number is such that the charge carried away by the protons and α -particles approaches the limit imposed by the atomic number of the nucleus, we would expect a decrease in fragment frequency.

3.2. *Velocity distribution of fragments*

Table 5 shows the average fragment velocity as a function of charge. Results are from both C2 and G5 emulsions. If we assume that the Coulomb repulsion between a splinter and the parent nucleus is roughly proportional to $Z(41 - Z)$, where 41 is an average atomic number for Ag and Br, and that the Coulomb barrier for α -particles is 10 MeV (Perkins 1949), then the velocity corresponding to the electrostatic repulsion is given by β_c .

TABLE 5

Z	2 (α 's)	3	4	5, 6	7, 8, 9 and 10
average obs. velocity	0.079	0.087	0.093	0.095	0.133
β_c	0.075	0.074	0.073	0.071	0.069
no. of fragments	—	15	23	16	7

Most of the fragments, therefore, have kinetic energies well in excess of the barrier repulsion. There is some evidence for a trend of increase of emission velocity with increasing charge, but it is difficult to be certain on this point, in view of the number of restrictions imposed in selecting the tracks.

One fragment each of Li, Be and B have been observed in the stars in G5 emulsions, all with ranges above 700μ , and velocities β of the order of 0.2 or more. Fragments of similar charge and velocity have been observed by Sørensen (1949). It is not possible to say whether these fragments form a 'tail' to the main distribution or are to be considered as generated in a quite different process from that responsible for the majority of the splinters.

3.3. *Angular distribution of fragments*

For C2 plates, the angle θ which the projection of a fragment in the plane of the emulsion makes with the downward vertical has been determined. These plates were exposed vertically at 11,000 ft., where the star-producing radiation is fairly well collimated. θ therefore measures effectively the projected angle which a fragment makes with the incident particle. The G5 plates were exposed at 80,000 ft., where the star-producing radiation has a very wide zenith-angle distribution. However, 75 % of the large stars are produced by ionizing particles, so that in most cases, θ can again be determined. The distribution for α -particles (all velocities), fragments of velocity $\beta < 0.14$, and $\beta > 0.14$ are given in figure 7. The α -particle distribution is almost isotropic. On the contrary, that for fragments of $\beta < 0.14$ is very strongly 'sideways' with respect to the incident particle. There is a fractional downward excess of only 21 ± 13 %; whilst the high-velocity fragments, especially those known to be of $\beta > 0.2$, are strongly collimated in the direction of the star-producing radiation. In what follows, we shall refer to this direction as 'downward'.

It is perhaps significant that the average downward component of velocity of the fragments in figure 7*b*, 0.02 ± 0.01 , is roughly equal to the downward velocity of the parent nucleus during the disintegration, also ~ 0.01 (Perkins 1950). As yet the evidence is insufficient for proof, but if these two facts are connected, then we could make the important conclusion that emission of fragments in this group takes place from the moving nucleus *after* it has received its full energy transfer. This hypothesis does not of course hold for the high-velocity group (figure 7*c*).

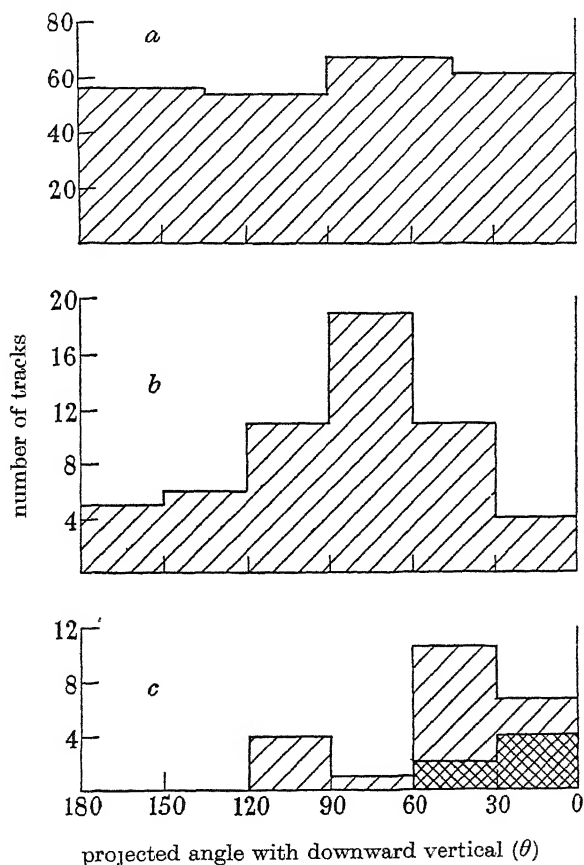


FIGURE 7. Projected zenith-angle distributions of fragments from stars. (a) α -particles (practically isotropic), (b) heavier fragments of velocity $\beta < 0.14$, (c) fragments of $\beta > 0.14$; \otimes fragments of $\beta > 0.2$.

3.4. Splinters and meson showers

From the data in G 5 plates, we can decide whether or not there is any connexion between production of star splinters and that of meson showers. We divide the stars into two groups, characterized by the numbers of shower particles present:

no. of shower particles per star	≤ 3 (mean = 1.4)	≥ 4 (mean = 7.1)
proportion of stars showing fragments	0.20 ± 0.05	0.25 ± 0.07

There is thus no significant difference in the splinter frequency in stars containing large or small numbers of shower particles.

3.5. *Multiple splinters*

The observed frequencies of emission of different numbers of splinters in a single star are as follows. Results are taken from both C2 and G5 plates.

total no. of stars	350
total no. of splinters	78
no. of stars with 1 splinter	55
no. of stars with 2 splinters	10
no. of stars with 3 splinters	1

In order to be identified, a splinter must have an angle of dip of less than 30° and be fairly stable, that is it must leave the nucleus entire or as a narrow 'cone' of disintegration products. Thus, since many splinters will be too unstable to be observed, or have steep angles of dip in the emulsion, it is probable that the true ratio of 'double' to 'single' events is much greater than the observed value. In the ten two-splinter events, the pair of splinters are in eight cases directed to opposite sides of the vertical. From considerations of conservation of momentum, this may indicate that not more than two splinters (stable or otherwise) are usually emitted.

An interesting feature regarding double or triple splinter events is revealed when the velocities of fragments in a *single star* are compared:

star					
1. triple splinter	$\begin{cases} Z \\ \beta \end{cases}$	3	6	8	
		0.08	0.12	0.16	
2. double splinter	$\begin{cases} Z \\ \beta \end{cases}$	5	9	—	
		0.08	0.17	—	
3. double splinter	$\begin{cases} Z \\ \beta \end{cases}$	4	5	—	
		0.06	0.07	—	
4. double splinter	$\begin{cases} Z \\ \beta \end{cases}$	4	8	—	
		0.09	0.11	—	
5. double splinter	$\begin{cases} Z \\ \beta \end{cases}$	4	6	—	
		0.10	0.09	—	

There is a distinct tendency, therefore, in an event in which two or more splinters are emitted, for the splinter of higher charge Z to have higher velocity β . This is quite opposite to the effect one expects from conservation of momentum.

4. INTERPRETATION OF RESULTS

The conclusions which have been arrived at may be summarized as follows:

(i) At high excitation energies, comparable with the total nuclear binding energy, a silver or bromine nucleus has a high probability, of the order of unity, of emitting one or more fragments. A fraction of these are stable; the remainder disintegrate spontaneously.

(ii) These fragments generally have kinetic energies in excess of their energy of Coulomb repulsion from the parent nucleus.

(iii) There is a strong tendency for the fragment to be emitted sideways from the incident particle.

(iv) No particular correlation has been found between production of splinters and meson showers.

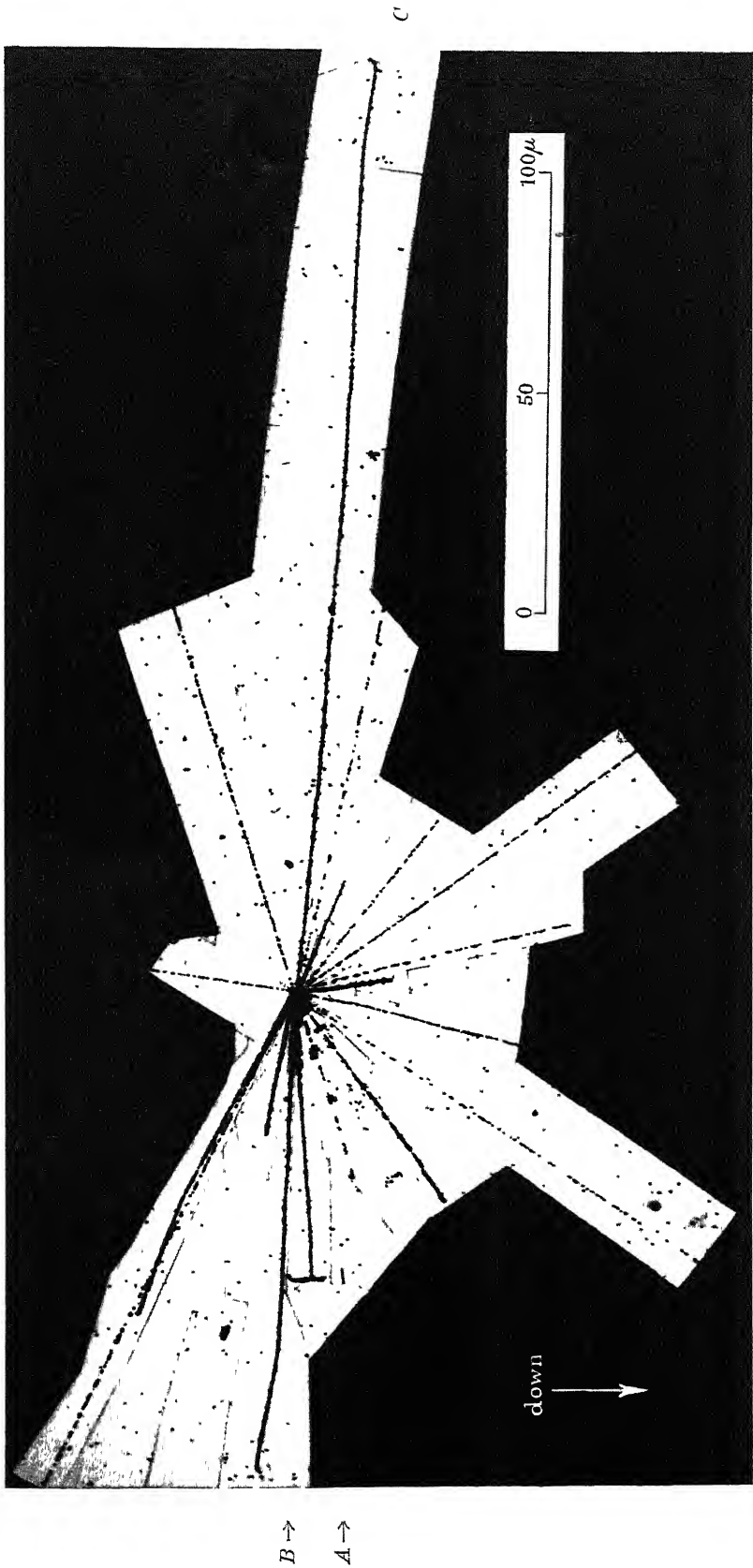


FIGURE 8. Production of three heavy splinters in a single star A, hammer track ($Z=3$),
B, fragment of $Z\sim 6$, C, fragment of $Z\sim 9$, 0.2 emulsion.

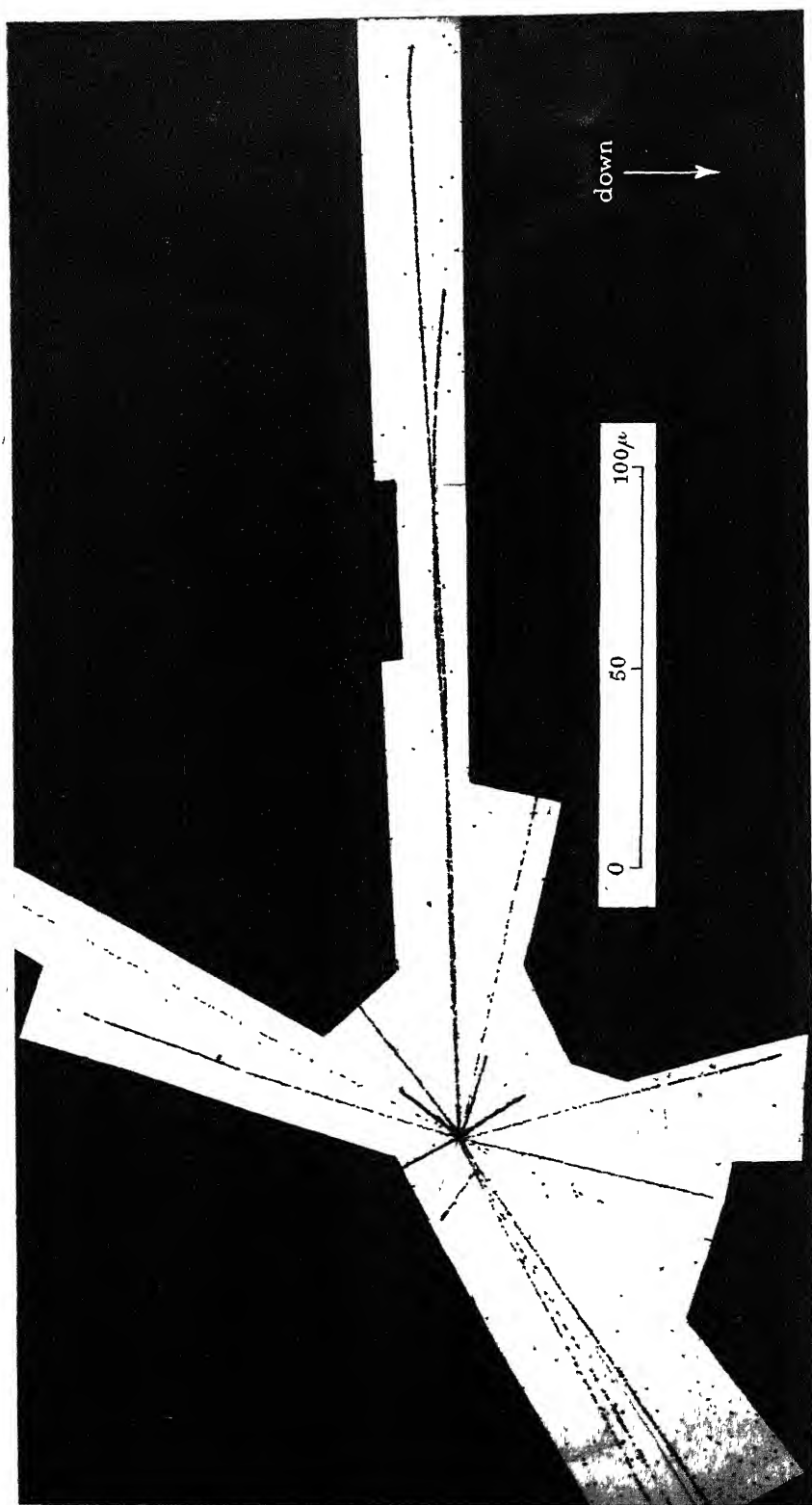


Figure 9. Emission in a nuclear explosion of a ^8Be nucleus in its ground state, with kinetic energy of 50 MeV. The ^8Be breaks up spontaneously into a pair of alpha-particles α with very small angle of separation. C2 emulsion.

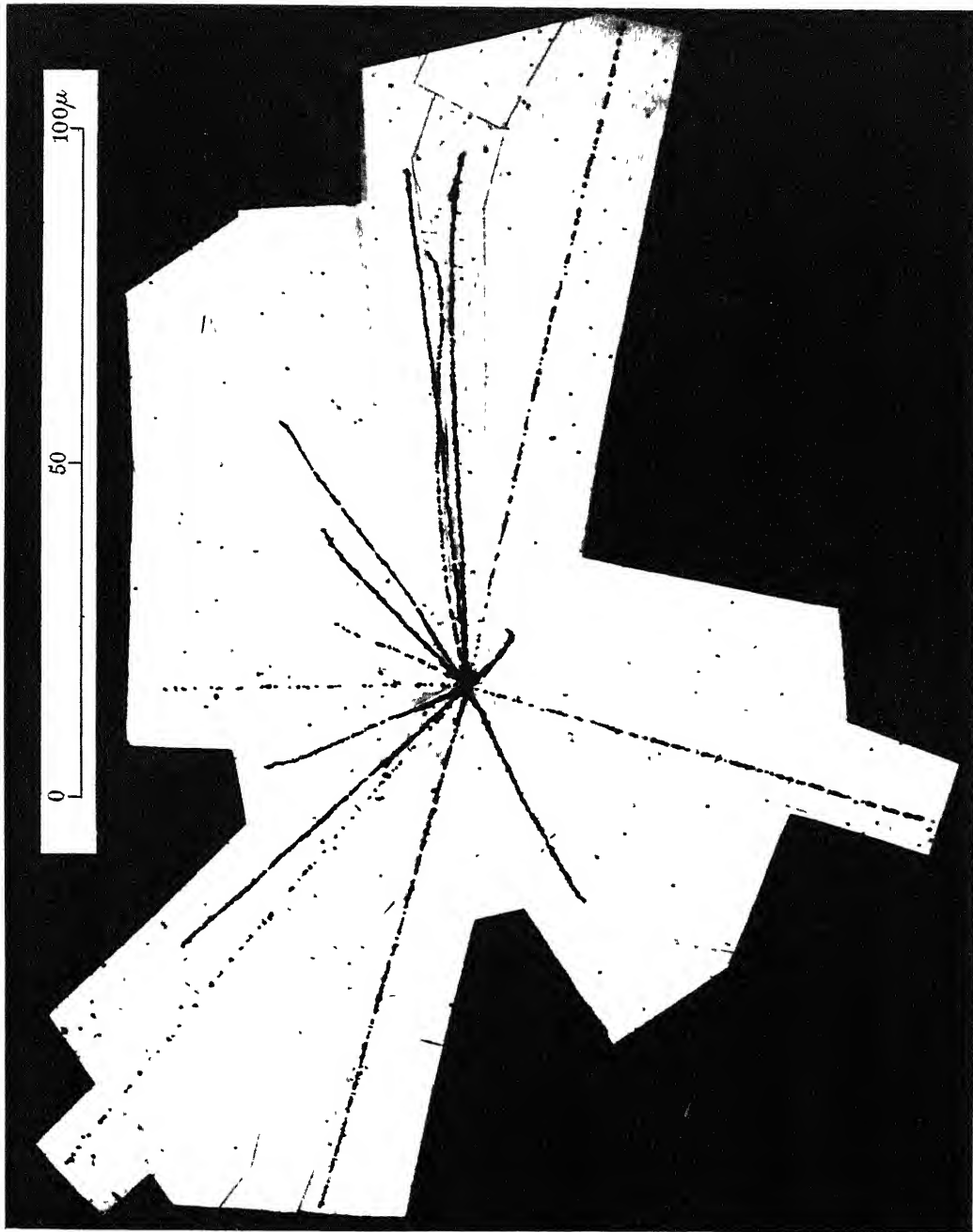


FIGURE 10. Photomicrograph of disintegration leading to emission of a ${}^9\text{B}$ nucleus A in its ground state. The fragment disintegrates into a proton and two alpha-particles of nearly equal range and velocity. C^2 emulsion.

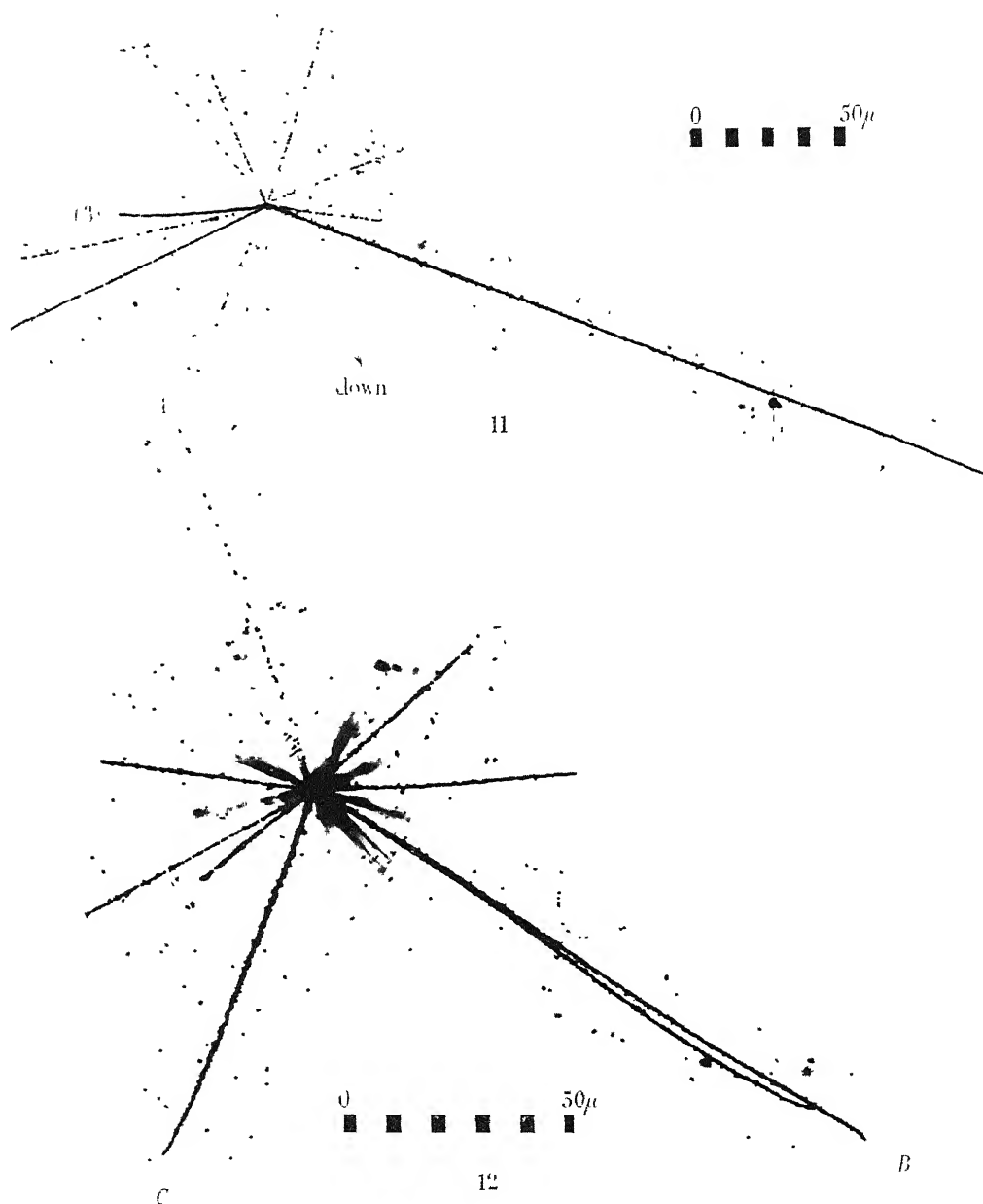


FIGURE 11. Two heavy splinters ejected in opposite directions in a single star. The charge of particle (3) is ~ 5 , particle (2) has charge ~ 9 and shows δ -rays along the track. (1) is the track of an alpha-particle. C2 emulsion.

FIGURE 12. Disintegration of silver or bromine nucleus by an incident proton A, leading to emission of a ${}^8\text{Be}$ α pair B and a heavy fragment C. The charge of C is greater than ten units, and the track shows characteristic taper. Many of the remaining tracks of the star have not been photographed. G5 emulsion.

- (v) The mean charge on a fragment does not depend much on the excitation energy of the nucleus. The average charge is ~ 5 , the average kinetic energy ~ 50 MeV.
- (vi) The average fragment velocity probably increases slightly with charge Z .

4.1. *The fission process*

The main features of the above phenomena can be understood, qualitatively at least, in terms of 'fission' occurring as a result of violent oscillations of the silver or bromine nucleus.

At low excitation energies U , as is well known, the 'star' phenomenon can be interpreted as a simple evaporation of the whole nucleus. For high values of U (> 500 MeV) it has been postulated (Harding, Lattimore & Perkins 1949) that the part of the nucleus struck by the incident particle may become so 'hot' as to result in a preliminary local evaporation. On the other hand, Tomonaga (1938) has shown that, even at high nuclear temperatures, the thermal conductivity of nuclear matter is very high, and Fujimoto & Yamaguchi (1950*a*) consider that local evaporation is unlikely. At the same time, the viscosity decreases very rapidly with increasing temperature, so that both surface and volume oscillations are possible for high U values.

The emission of fragments from a silver nucleus, as a result of purely surface oscillations, has been treated by Fujimoto & Yamaguchi (1950*a*), who point out that, owing to the decrease of surface tension with increase of temperature T , the fission width becomes comparable with the neutron width when $T \sim B$, the binding energy per nucleon. For $T > B$, fission should become a dominant process. This result would hold for symmetrical fission of a liquid drop, as in the (quasi-static) Bohr-Wheeler theory. In the present experiments, the fission is strongly asymmetric, but, as can be seen from table 4, the condition $T \sim B$ for the threshold holds as before. Assuming the energy U to be concentrated only in surface waves, their amplitude has been calculated by Bagge (1938), but a correction is required to account for the dependence of surface tension on nuclear temperature. The order of magnitude of mass of an emitted fragment is $a = (\overline{\Delta r^2})^{\frac{1}{2}}/r_0^3$, where $\sqrt{(\overline{\Delta r^2})}$ is the r.m.s. increase in nuclear radius, r_0 the nucleon radius. According to Yamaguchi (1950*b*), $a \sim 10$ for a heavy nucleus ($A = 100$) at high excitation energy. The number of such nucleon ensembles would be $n \sim (A/a)^{\frac{1}{2}} = 5$, and the energy of each would be $750/5 = 150$ MeV (taking the excitation energy equal to the nuclear binding energy). The average kinetic energy of a fragment (treating it as one of a number of simple-harmonic oscillators) would thus be ~ 75 MeV. These values are in fair agreement with the observations. However, such a thermodynamical model will only be strictly valid when U is less than the total nuclear binding energy. At the very excitation energies at which fission becomes a dominant process, a rigorous theoretical treatment will hardly be possible.

4.2. *Angular momentum considerations*

It has been suggested by Telegdi (1949) that the emission of fragments may be viewed as a process which rapidly dissipates angular momentum. The fragment may of course carry this off both as orbital momentum and as intrinsic spin. The incident particle will usually strike the nucleus to one side of the centre and therefore possess

considerable angular momentum with respect to it. It is, however, difficult to understand how the primary can transfer so much angular momentum to the nucleus that it cannot be dissipated by simple evaporation. It is worth noting that the angular momentum carried by a fragment will in general be considerably higher than that of a single emitted nucleon, and that if some of the fragments are built up of α -particle clusters, an anisotropic angular distribution might well result.

4.3. Long-range forces

One of the most significant results of the present investigation is that fragments are emitted with kinetic energies far in excess of their Coulomb repulsion. One has to explain, for example, how a nucleus of charge $Z = 8$ can acquire an energy of 200 MeV, when its electrostatic repulsion from the parent nucleus is only 50 MeV at most.

The absence of correlation between splinters and meson showers appears to preclude the hypothesis put forward by Heisenberg (1949). Heisenberg suggests that the 'turbulent effects' resulting from production of a meson shower, act as a 'gruel' which can drag out relatively large fragments from the parent nucleus. This is also not substantiated by the angular distribution of most of the splinters.

The forces responsible for fragment emission must clearly be stronger than the Coulomb forces. At the same time they must possess the long-range property of the latter, in order to act on an agglomerate of up to twenty nucleons and eject them with a very small internal energy, so that the fragment remains stable.

In conclusion, it may be remarked that the simple relativistic collision between two free nucleons has not yet been treated satisfactorily from a theoretical standpoint, an immediate solution of the corresponding problem of interaction between nucleon aggregates at high energy, is therefore hardly to be expected.

It is a pleasure to thank Professor C. F. Powell for extending to me facilities at the H. H. Wills Physical Laboratory, University of Bristol, during the later stages of the investigations; Dr E. P. George, of Birbeck College, University of London, for a generous supply of 'giant' stars in C2 plates; and Professor Sir George Thomson, for valuable discussions and advice.

[*Note added in proof.* Since this paper was communicated, J. Crussard (*C.R. Acad. Sci., Paris* 1950, **231**, 141) has published observations on the emission from stars of ^8Be in the ground state, in the form of α pairs. His results agree with those in § 2.2 (i).]

REFERENCES

- Bagge, E. 1938 *Ann. Phys., Lpz.*, **33**, 359.
 Bohr, N. 1941 *Phys. Rev.* **59**, 270.
 Bonetti, A. & Dilworth, C. 1949 *Phil. Mag.* **40**, 585.
 Freier, P., Lofgren, E., Ney, E. & Oppenheimer, F. 1948 *Phys. Rev.* **74**, 1818.
 Fujimoto, Y. & Yamaguchi, Y. 1950a *Progr. Theor. Phys. (Japan)*, **5**, 76.
 Fujimoto, Y. & Yamaguchi, Y. 1950b Private communication.
 Harding, J. B. 1949 *Phil. Mag.* **50**, 530.
 Harding, J. B., Lattimore, S. & Perkins, D. H. 1949 *Proc. Roy. Soc. A*, **196**, 325.
 Heisenberg, W. 1949 *Z. Phys.* **126**, 569.
 Hodgson, P. & Perkins, D. H. 1949 *Nature*, **163**, 439.

- Lovera, G. 1947 *Ric. Sci. Ricostruz.* **17**, 2045.
Occhialini, G. & Powell, C. F. 1947 *Nature*, **159**, 93.
Perkins, D. H. 1949 *Phil. Mag.* **40**, 601.
Perkins, D. H. 1950 *Phil. Mag.* **41**, 138.
Rosenfeld, L. 1948 *Nuclear forces*. Amsterdam: North Holland Publishing Company.
Schopper, E. 1947 *Naturwissenschaften*, **34**, 118.
Sorensen, S. O. C. 1949 *Phil. Mag.* **40**, 947.
Telegdi, V. L. 1949 Como Conference.
Tomonaga, S. 1938 *Z. Phys.* **110**, 573.
Wheeler, J. 1941 *Phys. Rev.* **59**, 27.
-

The derivation of a model solar chromosphere from radio data

By J. H. PIDDINGTON

Division of Radiophysics, C.S.I.R.O., Australia

*(Communicated by Sir Edward Appleton, F.R.S.—Received 13 July 1949—
Revised 20 March 1950)*

An empirical relationship between the equivalent black-body temperature of the sun's disk and the radio-frequency is derived from experimental results. The formula is applicable for frequencies between about 600 and 24,000 Mc./sec. An expression for the absorption coefficient of radio waves in the solar atmosphere is developed so that, by Kirchhoff's principle, it is possible to determine the emission from each level and hence the overall emission in terms of known or assumed conditions. The coronal and chromospheric components of radiation are separated and their distributions over the solar disk are determined theoretically. The conclusions are compared with experimental results and found to be in sufficiently good agreement for the present purposes.

From the reduced radio data an equation is derived relating the optical depth of a given level in the chromosphere with electron temperature at that level. Optical depth may be expressed as a function of electron temperature and density so that the equation may be used to check the validity of any proposed model chromosphere.

By combining the radio results with optical data in the form of intensities of spectrum lines at various levels a determination of electron density and temperature over a range of levels from about 5000 to 15,000 km. is made. The radio results are difficult to reconcile with Redman's estimate of an electron temperature of 30,000° K at 1500 km. A marked departure from conditions of hydrostatic equilibrium is indicated.

1. INTRODUCTION

Martyn (1946, 1948) and others (Pawsey & Yabsley 1949; Ginsburg 1946; Waldmeir & Müller 1948) have shown that most of the radio-frequency radiation from the quiet (relatively spot-free) sun is thermal. It seems safe to assume this over the frequency range (600 to 24,000 Mc./sec.) discussed below. Using the best available estimates of temperature and pressure, Martyn determined the absorption coefficient for radio waves throughout a range of levels in the solar atmosphere. The application of Kirchhoff's law then allowed the emission from each layer to be determined and hence the total emission for each of a series of annuli on the solar disk. At the lower frequencies, the corona was found to be a partial reflector, the refractive index departing sensibly from unity, and absorption occurring mainly at the turning point

of the rays, such cases will not concern us here. At the higher frequencies the optical depth τ becomes large before appreciable refraction occurs. Martyn's results give the distributions of intensity across the disk at each frequency, and of average intensity with frequency for the whole disk. The available experimental results are in general agreement with Martyn's predictions; limb brightening is found (Piddington & Minnett 1949*a*; Minnett & Labrum 1950, Piddington & Hindman 1949), as predicted, but the quantitative agreement between theory and observation is not close, partly because of the approximations used and partly because of insufficient knowledge of chromospheric pressures and temperatures.

In this paper the inverse of Martyn's problem is attempted, using recent measurements of average emission intensities over the solar disk we try to determine chromospheric conditions. To do so we must first find the distribution of chromospheric emission with frequency at the centre of the disk. For the (radio) frequencies considered, most of the emission comes from the chromosphere,* but there is a coronal component whose average intensity over the disk must be computed and subtracted from the data. The residue is the chromospheric emission averaged over the disk. From the known frequency distribution of this radiation it is shown that the distribution over the disk at each frequency may be determined and hence the (required) intensity distribution at the centre of the disk. The estimates of the distribution of total radiation over the disk which result incidentally are compared with experimental observations and found sufficiently good agreement for the present purposes.

Using the radio-frequency emission spectrum from the centre of the disk an integral equation may be set up relating observed intensity of emission to electron temperature and optical depth, both unknown functions of height above the photosphere. Fortunately, this equation can be solved uniquely and exactly so that we can relate electron temperature and optical depth. Since optical depth may be expressed as a function of electron temperature and density we have determined a relationship between these two quantities, dependent on radio data only. This does not provide height distributions of these quantities but provides criteria which any given model chromosphere must satisfy, and by combining it with certain optical data, electron densities and temperatures throughout a range of the chromosphere are found.

It is also shown that the radio data are difficult to reconcile with Redman's (1942) estimate of an electron temperature of about 3×10^4 °K at a level of 1500 km. They also indicate a departure from conditions of hydrostatic equilibrium.

2. THE OBSERVED INTENSITIES OF EMISSION

The intensity of radio-frequency emission of a black body is (within the temperature limits concerned here) linearly related to the temperature according to the Rayleigh-Jeans formula, so that equivalent black-body temperature may be used as a measure of radiation intensity. Its use is justified for certain radio-frequency problems because of the simple physical relationship between gas temperature and optical depth and equivalent temperature. It also provides a simple unit of intensity of

* This region is defined here as extending from a few hundred to about 15,000 km. above the photosphere.

emission or 'source brightness'. Eclipse observations (Minnett & Labrum 1950; Piddington *et al.* 1949) indicate that at the frequencies under consideration most of the observed radiation originates on the visible disk of the sun. We will assume that all radio emission discussed below takes place from the visible disk, so that the average equivalent temperature of the disk is a measure of the flux density of radiation received on the earth.

Measurements of flux density expressed as equivalent disk temperatures of the quiet sun have been made at frequencies between 600 and 24,000 Mc./sec. These results, together with the source of reference, are listed in table 1.

TABLE 1

frequency (Mc./sec.)	minimum equivalent disk temperature (° K $\times 10^4$)	observer and reference
24,000	1.0-1.1	Dicke & Beringer (1946)
24,000	1.0	Piddington & Minnett (1949 <i>a</i>)
9,400	2.3	Sander (1947)
9,400	1.7	Minnett & Labrum (1950)
3,000	5.4	Piddington & Hindman (1949)
2,800	5.6	Covington (1947)
1,200	20	see subscript*
600	50	Lehany & Yabsley (1949)

* Measurements by Lehany & Yabsley (1949) gave a value 1.0×10^5 °K. More recent unpublished observations by the author and H. C. Minnett indicated a value of at least 2.0×10^5 °K. In a private communication Mr Yabsley has informed the author of an experimental error in the early results which should be revised to about 2.0×10^5 °K.

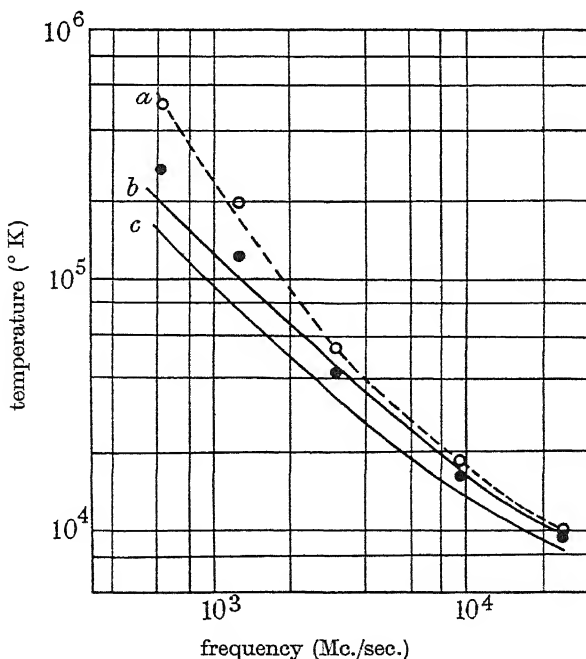


FIGURE 1. Equivalent disk temperature plotted against frequency. *a*, observed equivalent disk temperatures. *b*, temperatures corrected for coronal contribution. *c*, temperatures at the centre of the disk.

The more recent measurements (Minnett & Labrum 1950) at 9400 Mc./sec. are thought to represent quieter conditions and are accepted rather than the somewhat higher value found by Sander. The reference source is marked against each point, the 1200 Mc./sec. value being marked with an asterisk. The equivalent disk temperatures plotted against frequency are indicated by circles in figure 1. The dashed curve *a* is drawn through the points.

3. THE ABSORPTION COEFFICIENT OF RADIO WAVES IN THE SOLAR ATMOSPHERE

Under certain conditions, which are discussed in appendix 1, the absorption coefficient of electromagnetic waves in an ionized atmosphere is given by the well-known expression, due to Lorentz,

$$\kappa = \frac{N\epsilon^2\nu}{\pi cmf^2}, \quad (1)$$

where *N*, *ε*, *m*, *ν* are electron density, charge (e.s.u.), mass and collision frequency, *f* is the wave frequency and *c* the velocity of light. For the frequencies considered and the appropriate levels the equation is found to be valid.

In determining *ν*, the presence of neutral hydrogen atoms is neglected, the question being discussed in appendix 1.

Martyn (1948) has shown that an expression for *ν* may be found from gas-diffusion theory. We find a slightly different (numerically) expression derived by Chapman & Cowling (1939, pp. 177 and 333):

$$\nu = \frac{\sqrt{\pi} N \epsilon^4}{\sqrt{(2m)} (kT_e)^{\frac{3}{2}}} \log_e \left\{ 1 + \left(\frac{4kT_e}{\epsilon N^{\frac{1}{2}}} \right)^2 \right\},$$

where *T_e* is the electron temperature, *k* is Boltzmann's constant and the velocity distribution is Maxwellian. Substituting in equation (1) and inserting numerical values

$$\kappa = \frac{\zeta N^2}{f^2 T_e^{\frac{3}{2}}}, \quad (2)$$

where *ζ* is a slowly varying function of *T_e* and *N* given below. A similar expression for *κ* may be derived from Gaunt's (1930) quantum-mechanical treatment of radiation by free-free transition of an electron in the field of a proton, *ζ* being a slowly varying function of *T_e* and *f*. Other values of *ζ* have been derived by Townes (1947) using Kramer's classical derivation of continuous X-ray emission produced by bombarding nuclei with electrons and by Henyey & Keenan (1940) using quantum-mechanical methods. Values of *ζ* calculated from the various formulae and for several conditions of temperature, pressure and frequency are listed in table 2.

Values of 0.07 and 0.10 will be adopted for *ζ* in the chromosphere and corona respectively.

TABLE 2

derivation	value of ζ		
	$N = 1 \text{ cm.}^{-3}$ $T = 10^4 \text{ }^\circ \text{K}$ $f = 1.8 \times 10^8 \text{ sec.}^{-1}$	$N = 10^9 \text{ cm.}^{-3}$ $T = 10^5 \text{ }^\circ \text{K}$ $f = 10^9 \text{ sec.}^{-1}$	$N = 10^{10} \text{ cm.}^{-3}$ $T = 10^4 \text{ }^\circ \text{K}$ $f = 10^{10} \text{ sec.}^{-1}$
Gaunt	0.14	0.14	0.10
Townes	0.12	0.10	0.07
Henney & Keenan	0.10	—	—
present estimate	0.12	0.09	0.07

4. THE CORONAL CONTRIBUTION TO THE EQUIVALENT DISK TEMPERATURE

The coronal contribution to the equivalent disk temperature is found approximately by considering the radiation from that half of the corona nearest the earth. This neglects a volume in the form of an annulus lying behind the sun (but visible from the earth), the cross-section of which is shown as area A of figure 2. The corona

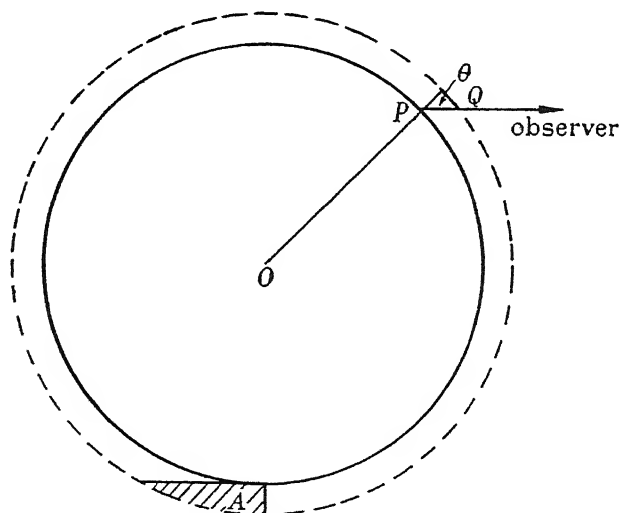


FIGURE 2. Radiation from the corona.

is also assumed to be optically thin even near the limb of the sun. The errors involved are not important for our present purposes as will be seen below.

Consider an elementary volume of the corona in the form of a hemispherical shell concentric with the sun, of radius η times the solar radius and thickness $ds \doteq 7 \times 10^{10} d\eta \text{ cm.}$ The surface area of the shell is $2\eta^2$ times that of the solar disk, so that its contribution to equivalent disk temperature is $2\eta^2$ times that of a flat shell of thickness ds covering the visible disk, that is, $2\eta^2 \kappa T_e ds$. The total contribution by the corona to the equivalent disk temperature is found by integration over suitable limits:

$$T_c = \int 2\eta^2 \kappa T_e ds.$$

Let T_e have the accepted value of $10^6 \text{ }^\circ \text{K}$ and N be given by Allen's (1947) empirical formula:

$$N = 10^8 (1.55\eta^{-6} + 2.99\eta^{-16}) \text{ cm.}^{-3}.$$

The absorption coefficient κ is given by equation (2), ζ having a value of 0.1. Substituting these values and integrating between the limits $\eta = 1.02$ (a level of 14,000 km.) and ∞ we find

$$T_c = \frac{1.02 \times 10^{23}}{f^2}. \quad (3)$$

It will be found desirable later to have an estimate of the distribution of the coronal component over the disk. The region from which appreciable radiation is received is a shell of thickness only a fraction of the solar radius. The thickness of this shell when viewed at an angle θ to the solar radius (along the line PQ of figure 2) is approximately proportional to $\sec \theta$. The distribution of radiation may, therefore, be represented approximately by a $\sec \theta$ law, and the coronal contribution from a region defined by the angle θ is given by

$$T_c(\theta) = \frac{5.1 \times 10^{22}}{f^2} \sec \theta. \quad (4)$$

The numerical constant in this equation is chosen so that the average value of $T_c(\theta)$ over the disk is equal to T_c given by equation (3).

The distribution represented by equation (4) is reasonably accurate over a large part of the disk but obviously does not hold near the limb where an infinite value of T_c would be indicated. The discrepancy is due to a neglect of self-absorption which must limit the equivalent temperature to the electron temperature and also to the geometrical error involved in using a $\sec \theta$ law near the limb. However, we are concerned primarily with the total coronal contribution and not its detailed distribution, and this, as will be seen later, is probably given by (3) with sufficient accuracy.

Values of T_c calculated from equation (3) are given for the various frequencies in table 3. These are subtracted from the observed values of equivalent disk temperature to give the chromospheric component (T_D) which is listed in the fourth column. In calculating T_D for 600 Mc./sec. an allowance has been made for the absorption of chromospheric radiation by the corona which has an appreciable optical depth at

TABLE 3

frequency (Mc./sec.)	observed disk temperature (° K)	T_c (° K)	chromospheric component of equivalent disk temperature T_D (° K)
24,000	10,000	180	9,800
9,400	17,000	1,150	15,800
3,000	54,000	11,400	42,600
1,200	200,000	71,000	129,000
600	500,000	284,000	285,000

this frequency. This has been done as follows: equation (3) gives the coronal contribution to the average equivalent disk temperature, which, if the optical depth is taken as small, is equal to the product of the electron temperature (a constant 10^6 ° K) and the average optical depth. Hence the average value of τ is 0.28, and

chromospheric radiation is reduced by a factor $e^{0.28} = 1.32$. Thus the true corrected disk temperature is given by $(500,000 - 284,000) 1.32$ or 285,000. The above method of treatment has evidently begun to break down for frequencies near 600 Mc./sec., but the results are approximately correct while at higher frequencies re-absorption is negligible.

An empirical expression for T_D which gives values approximating those in table 3 for the three higher frequencies is

$$T_D = 4800 + 1.2 \times 10^{14} f^{-1}. \quad (5)$$

This is shown as curve *b* of figure 1, while the values of T_D from table 3 are marked as filled circles. The curve is a close fit at the higher frequencies but falls somewhat below the lower frequency points.

5. THE DISTRIBUTION OF CHROMOSPHERIC RADIATION OVER THE DISK

Consider a thin spherical chromospheric shell concentric with the sun and of thickness dx . The optical depth when viewed in a direction making an angle θ with the solar radius (the line PQ of figure 2) is

$$d\tau = \kappa \sec \theta dx,$$

which, with equation (2), becomes

$$d\tau = \frac{\xi N^2 dx}{T_e^{\frac{1}{2}} f^2 \cos \theta}. \quad (6)$$

The equivalent temperature of the chromosphere when viewed along PQ is given by (see, for example, Piddington & Minnett 1949*b*)

$$T = \int T_e e^{-\tau} d\tau, \quad (7)$$

integration being throughout the chromosphere. If, for the present purpose, we assume the chromosphere to be horizontally stratified so that N and T_e are functions of x only, then equation (7) may be written

$$T(f, \theta) = \int F\{T_e(x), \tau(f, \theta, x)\} dx,$$

where F is a function and $T_e(x)$ and $\tau(f, \theta, x)$ are the values of T_e and τ at a level x , frequency f , and angle of emergence θ . But, from equation (6), f and θ are always combined in the term $f^2 \cos \theta$, so that we may write

$$T(f, \theta) = \int F\{T_e(x), \tau_0(f \cos^{\frac{1}{2}} \theta, x)\} dx,$$

where τ_0 is the value of τ when $\theta = 0$ and is not an independent function of both f and θ but only of $f \cos^{\frac{1}{2}} \theta$.

It follows, therefore, that $T(f, \theta) = T_0(f \cos^{\frac{1}{2}} \theta), \quad (8)$

where T_0 is the value of T when $\theta = 0$. This means that the equivalent temperature at frequency f for a ray emerging at an angle θ is the same as that at frequency $f \cos^{\frac{1}{2}} \theta$ for a ray from the centre of the disk.*

Now T_D may be expressed in terms of $T(f, \theta)$ by integrating the latter over the disk:

$$\begin{aligned} T_D &= \frac{1}{\pi} \int_0^{\frac{1}{2}\pi} 2\pi \sin \theta \cos \theta T(f, \theta) d\theta \\ &= 2 \int_0^{\frac{1}{2}\pi} \sin \theta \cos \theta T_0(f \cos^{\frac{1}{2}} \theta) d\theta. \end{aligned}$$

This integral equation may be solved by writing

$$q = f \cos^{\frac{1}{2}} \theta,$$

$$dq = -\frac{1}{2} f \sin \theta \cos^{-\frac{1}{2}} \theta d\theta.$$

The equation becomes

$$T_D = 4 \int_0^f \frac{x^3}{f^4} T_0(x) dx,$$

where f is constant under the integral sign. Differentiating with regard to f

$$\frac{d}{df} [f^4 T_D] = 4f^3 T_0(f),$$

which reduces to (9)

$$f \frac{d}{df} T_D + 4T_D = 4T_0(f).$$

Substituting for T_D from equation (5) we have

$$T_0 = 4800 + 9.1 \times 10^{13} f^{-1}. \quad (10)$$

Equation (9) is shown plotted as curve *c* in figure 1.

The data contained in equation (10) is all that is needed to find the required relationship between electron temperature and optical depth. It appears desirable, however, to digress for a moment to find an expression for the distribution of the total intensity of radiation over the disk. Such an expression is of some interest itself and also allows a rough check on the above theory by comparison with experimental data. Equations (4), (8) and (10) allow us to determine the equivalent temperature at any point on the solar disk at any frequency. Combining them we have the total radiation (including the coronal component) from any point on the disk and for any frequency

$$T_t(f, \theta) = 4800 + \frac{9.1 \times 10^{13}}{f \cos^{\frac{1}{2}} \theta} + \frac{5.12 \times 10^{22}}{f^2 \cos \theta}, \quad (11)$$

where the first two terms are chromospheric and the last due to the corona. The equation does not, as we have seen, apply near the limb of the sun.

* A corollary to this theorem is as follows: Within the limits where equation (1) holds, no more information can be found by measuring the equivalent temperature for various values of θ and a fixed frequency than by measuring the equivalent temperature at the centre of the disk over a range of frequencies. The reverse applies in the optical spectrum.

In appendix 2 the distribution of intensity given by equation (11) is compared with experimental data. The comparison which involves estimates of 'limb brightening' is rather arbitrary but provides a rough check on the theory leading to (11). Agreement is generally as good as the method of comparison would warrant.

6. A RELATIONSHIP BETWEEN ELECTRON TEMPERATURE AND DENSITY IN THE CHROMOSPHERE

Equation (7) may be applied to radiation from the chromosphere at the centre of the disk. The upper limit of integration is the top of the chromosphere, but this may be extended to infinity with only a small error as will be seen later. Thus we have

$$T_0 = \int_0^\infty T_e e^{-\tau} d\tau, \quad (12)$$

where

$$T_0 = a + b/f,$$

a and b being given in equation (10).

Equation (12) may be solved* by writing $\tau = \alpha/f^2$ and multiplying both sides by f^2 :

$$\alpha f^2 + b f = \int_0^\infty T_e e^{-\alpha/f^2} d\alpha,$$

where T_e is a function of α so that the right-hand side is the Laplace transform of T_e . From tables of Laplace transforms we find

$$T_e = a + \frac{b}{\Gamma(\frac{1}{2}) \alpha^{\frac{1}{2}}} = a + \frac{b}{\sqrt{(\pi \alpha)}}$$

or

$$T_e = a + \frac{b}{\sqrt{(\pi f^2 \tau)}}. \quad (13)$$

Thus, from the radio data we have a relationship between electron temperature and optical depth in the chromosphere. Also, since $\tau = \int_y^\infty \kappa dx$, we have from equation (2)

$$\alpha = 0.07 \int_x^\infty \frac{N^2}{T_e^{\frac{3}{2}}} dx.$$

Substituting the values of a and b from equation (9)

$$T_e = 4800 + 1.94 \times 10^{14} \left\{ \int_y^\infty \frac{N^2}{T_e^{\frac{3}{2}}} dy \right\}^{-\frac{1}{2}}, \quad (14)$$

which is a relationship between electron temperature and density in the chromosphere derived from radio observations. The range of values of T_e for which (14) is valid will correspond approximately to the range of equivalent radio temperatures defined by equation (10), namely, 8600°K to $1.6 \times 10^5^\circ \text{K}$.

While equation (13) is not alone sufficient to determine electron temperature and density at all levels, it may be used to test any model of the chromosphere or, if

* Thanks are due to Dr R. A. Sack for this method of solution.

another relationship between N and T_e is known, then a solution may be found. In the following section the results of observations of the intensity of certain spectrum lines emitted by the chromosphere were used to determine such a relationship.

7. SPECTRUM-LINE INTENSITIES

Menzel & Cillié (1937) have shown that ionized hydrogen in which the velocity distribution is Maxwellian emits in the Balmer continuum according to the equation

$$E_c = K_1 \frac{N^2}{T_e^{\frac{3}{2}}} \exp[(\chi_2 - hu)/kT_e], \quad (15)$$

where E_c is the energy emitted per unit volume per second, per unit wave-number interval, K_1 is a constant, depending only on atomic constants, χ_2 is the ionization energy of hydrogen in the second quantum state, hu is the energy of an emitted quantum, and k is Boltzman's constant. N and T_e are as defined above. Near the head of the continuum $(\chi_2 - hu)$ is small so that

$$E_c \propto \frac{N^2}{T_e^{\frac{3}{2}}}. \quad (16)$$

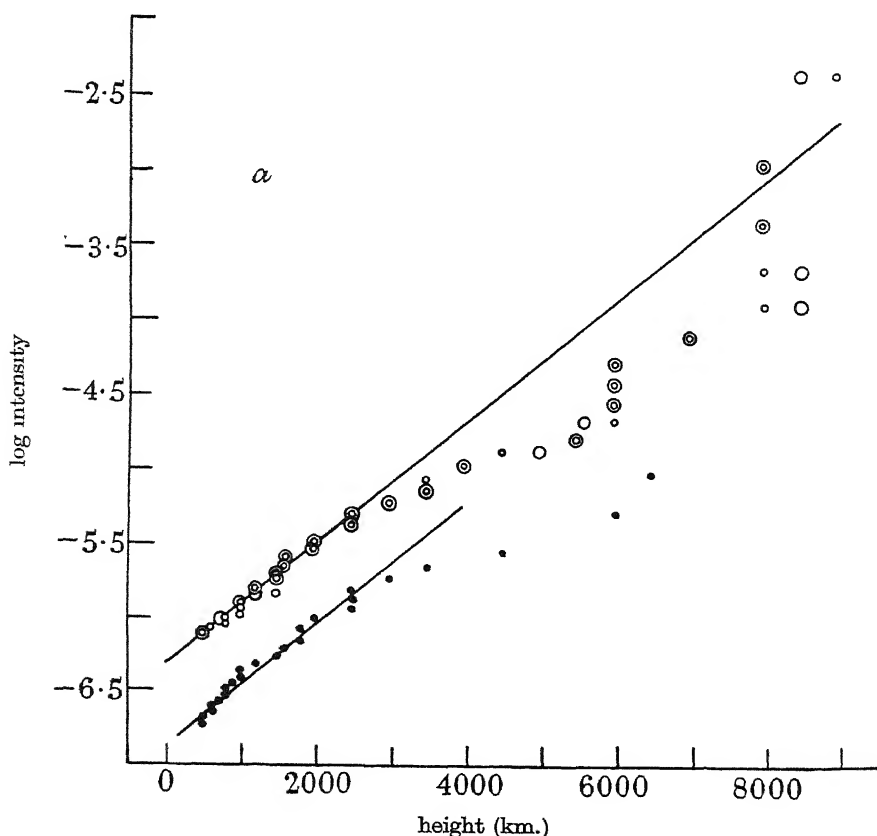


FIGURE 3. Visual light emission from the chromosphere.

a, log intensity-height diagram of the Balmer series (open circles) and Paschen series (filled circles); the larger open circles refer to the heights estimated from the eclipse spectra of 1905 and 1925, and the smaller ones to the eclipses of 1930 and 1937.

A similar relationship to (15) holds for line emission, and Menzel & Cillié have shown experimentally that for lines of high quantum number it is approximately true over a considerable portion of the chromosphere, where thermodynamic equilibrium does not exist. We will assume that in that part of the chromosphere with which we are concerned emission in lines or continua is proportional to $N^2/T_e^{\frac{1}{2}}$. It may be noted that the number of collisions/cm.³/sec. of electrons with protons, as derived in §3, is proportional to $N^2/T_e^{\frac{1}{2}}$. Evidently, the number of 'collisions' influencing radio-wave propagation is (as might be expected) proportional to the number resulting in light emission, the constant of proportionality involving only atomic constants.

Using observations of intensity in the Balmer continuum at levels of 670 and 1700 km. above the base of the chromosphere, Cillié & Menzel (1935) find that the value of $N^2/T_e^{\frac{1}{2}}$ may be expressed as

$$L = L_0 e^{-2\beta z}, \quad (17)$$

where

$$L_0 = 1.45 \times 10^{17},$$

$$\beta = 0.77 \times 10^{-8} \text{ cm.}^{-1},$$

and z is the height in cm. above the base of the chromosphere. By using observations of Balmer lines this expression for L was found to hold up to levels of 3170 km. The

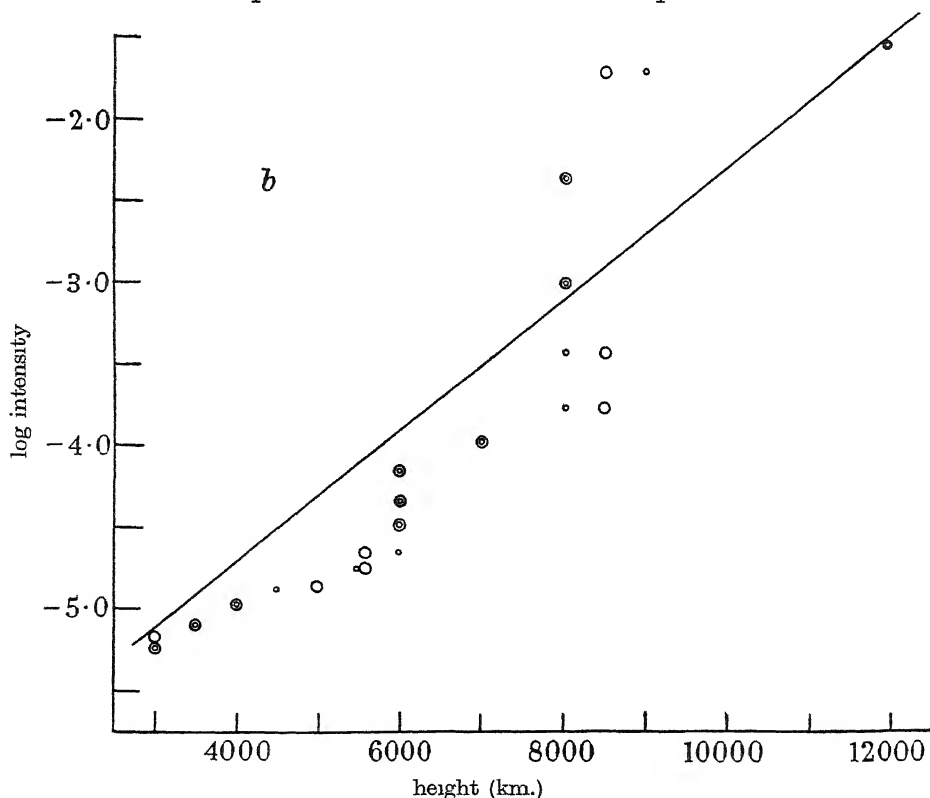


FIGURE 3b. Log intensity-height diagram of the Balmer series, based on the same observations as figure 1a. The intensity logarithms of figure 1a have been corrected by adding the empirical colour corrections. (Reproduced from *Astrophys. J.* 1947, 105, 55.)

result is based on an assumed exponential decay of L which is in reasonable agreement with the author's subsequent conclusions.

Using heights to which hydrogen lines are observed as data, Wildt (1947) has plotted log intensity against height for a number of lines of the Balmer and Paschen series. The observations range between levels of about 500 and 12,000 km. Wildt assumes a constant temperature in the chromosphere and concludes* that the density gradient up to a level of 15,000 km. is given by $\exp[-0.46 \times 10^{-8}z]$. Although our results indicate the existence of a marked temperature gradient in the chromosphere, thereby disagreeing with Wildt's conclusions, his plots of intensity against height will be used to interpret the radio results. The relevant data are reproduced in figure 3*a* (500 to 9000 km.) and figure 3*b* (3000 to 12,000 km.). The lower line in figure 3*a* refers to the Paschen series. Wildt concludes that the intensity gradient corresponds to a value of $\beta = 0.46 \times 10^{-8} \text{ cm.}^{-1}$, and that at least some of the deviations from the line drawn with this slope (figure 3) are due to accidental errors in the estimated heights.

Some of the points plotted by Wildt correspond to lines of low quantum number and the value of L should be corrected to correspond to the correct value of electron temperature. In most cases the corrections are small, but a few of the points at the higher levels should have increased values of L . The application of these corrections reduces the scatter of points.

It will be seen that the value of β found by Wildt is much lower than that of Cillié & Menzel (over a restricted region). Mitchell (1930) has observed that the gradient may change considerably from one eclipse to the next. The value chosen for the ensuing analysis is not very important, as it only varies the scale height of the model chromosphere without appreciably affecting the relationship between N and T_e . Analyses for β between 0.4×10^{-8} and $0.8 \times 10^{-8} \text{ cm.}^{-1}$ have been carried out, and we adopt a value of $\beta = 0.60 \times 10^{-8} \text{ cm.}^{-1}$ above a level of 1500 km. Below 1500 km. we use Cillié & Menzel's value of β , namely, $0.77 \times 10^{-8} \text{ cm.}^{-1}$. Thus from equation (17) we have

$$L = 1.44 \times 10^{16} \exp[-1.2 \times 10^{-8}x], \quad (18)$$

where x is the height in cm. above the 1500 km. level. This expression for $L = N^2/T_e^{\frac{3}{2}}$ holds for values of x between zero and 10,500 km., or further if it is agreed to extrapolate Wildt's results.

8. A MODEL CHROMOSPHERE

If the value of $N^2/T_e^{\frac{3}{2}}$ given by equation (18) is substituted in equation (14) values of N and T_e throughout a range of levels are determined. First, however, it should be noted that the integration in equation (14) is over the full range of levels from which radiation is received, while equation (18) is only valid between the levels indicated. Even at the highest frequencies no radiation is received from levels as low as 1500 km. because of the very large optical depth. When the coronal component was subtracted from the total observed radiation it was assumed that no chromospheric radiation was emitted from levels above 14,000 km. If equation (18) holds

* This conclusion is, apparently, misquoted in Wildt's paper where β is given as $0.92 \times 10^{-8} \text{ cm.}^{-1}$.

to infinite values of x an error will be involved which may be measured by calculating the optical depth to a level of 14,000 km. At the lowest frequency concerned this is found to be about 0.1, which is not serious but again indicates that the theory is becoming inaccurate near 600 Mc./sec. We conclude that equation (18) may be used safely for the full range of levels from which radiation is received.

Solving (14) and (18) we find

$$T_e = 4800 + 178 \exp [0.60 \times 10^{-8}x]. \quad (19)$$

Again, substituting this expression for T_e in equation (18) we have

$$N = 1.20 \times 10^8 (4800 + 178 \exp [0.60 \times 10^{-8}x])^{\frac{1}{2}} \exp [-0.60 \times 10^{-8}x]. \quad (20)$$

The distributions of N and T_e given by (19) and (20) are given in figure 4, the heights shown being above the base of the chromosphere. Above a level of 12,000 km. the curves depend on an extrapolation of Wildt's data, while below about 5000 km. they depend on an extrapolation of the radio data.

9. OTHER DETERMINATIONS OF ELECTRON DENSITY AND TEMPERATURE

It is of interest to compare the above estimates of electron density and temperature with those determined by other means. The electron density curve at high levels could merge with that of Baumbach (1937), plotted as curve *c* of figure 4 for levels above 14,000 km. Radio observations at 200 Mc./sec., where nearly all the radiation originates in the corona, indicate (Pawsey & Yabsley 1949) an equivalent temperature of about 10^6 ° K. This corresponds with the best estimates of coronal temperature, but the level at which such a temperature is reached is not indicated by the radio data.

At lower levels the radio results may be considered to hold to an approximate level of 5000 km., where the optical depth at the highest frequency (24,000 Mc./sec.) is unity. Spectroscopic determinations of electron temperature based on relative intensities of spectrum lines and of different portions of a continuum refer to lower levels. Thus Wildt (1947) has corrected estimates of Menzel & Cillié to find temperatures of 5240 and 4010° K at 670 and 900 km. respectively. Goldberg (1939) found temperatures of 4300 and 6700° K at 670 and 2330 km. respectively. Petrie's (1944) results are in reasonable agreement. All these results lie close to an extrapolation of the temperature curve of figure 4. Redman (1942), however, from observations of line broadening thought to be due to Doppler shift, found a gas temperature of 30,000° K at a level of 1500 km., a value which it is very difficult to reconcile with the electron-temperature curve of figure 4.

A more direct comparison of Redman's result with the radio data can be made by using the 24,000 Mc./sec. results of Piddington & Minnett (1949*a*). The equivalent disk temperature was 10,000° K, with an error of less than 5 % and fluctuations less than 5 % even during very disturbed conditions. 'Limb brightening' was also detected, so that the equivalent temperature at the centre of the disk could safely be taken as less than 10,000° K.

Now let us assume for the moment that the gas temperature at the 1500 km. level is 30,000° K. Using Cillié & Menzel's (1935) observations of the intensity of emission

in the Balmer continuum and a value of $T_e = 30,000^\circ \text{K}$, we find from equation (18) an electron density of $2.7 \times 10^{11} \text{ cm}^{-3}$ at this level. The absorption coefficient for 24,000 Mc./sec. radiation is given by putting $\zeta = 0.07$ in equation (2):

$$\kappa = 1.7 \times 10^{-6},$$

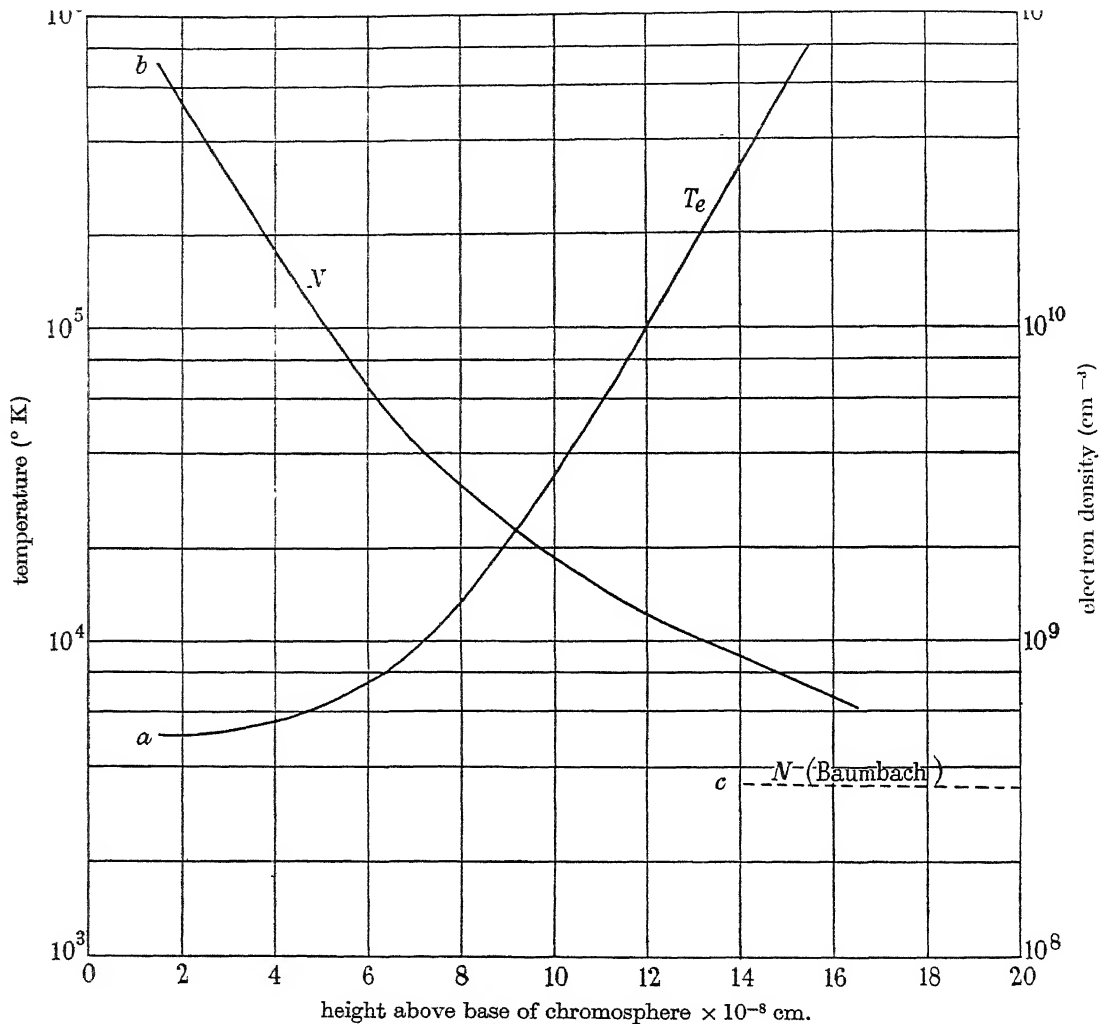


FIGURE 4. Electron densities and temperatures in the chromosphere. *a*, electron temperatures from radio data. *b*, electron densities from radio data. *c*, electron densities after Baumbach.

so that a layer only 10 km. thick would have an optical depth 1.7 and would provide nearly full black-body radiation. There can be no question of reduced emission caused by a rapid decrease in the dielectric constant of the medium in the manner discussed by Martyn (1948), since the electron density, even at the 1500 km. level, is too low. Also if the temperature at 1500 km. is less than $30,000^\circ \text{K}$ absorption will be even more rapid. We conclude that all 24,000 Mc./sec. radiation emerging from the sun's atmosphere does so from levels above 1500 km., and that the average electron temperature in the region from which emission takes place is about $10,000^\circ \text{K}$.

This conclusion and Redman's result could be reconciled if a temperature maximum existed between the photosphere and a level of about 5000 km. Such a maximum, reaching the large value of $30,000^\circ\text{K}$., would appear very improbable, however. Another explanation of the conflicting results might be sought in terms of localized regions of high temperature (greater than $30,000^\circ\text{K}$) extending outwards from about 1000 km. above the photosphere. If such regions were present during the observations of Piddington & Minnett their total area must have been constant to about 2 % of the area of the solar disk, otherwise they would have been detected. An explanation of Redman's results in terms of 'hot areas' would appear unlikely.

10. DEPARTURE FROM CONDITIONS OF HYDROSTATIC EQUILIBRIUM

It has long been thought that a supporting force of some kind opposes gravity in the chromosphere. The curves of figure 4 allow the model chromosphere to be tested for the presence or otherwise of hydrostatic equilibrium. At the lower levels (below that at which the temperature has reached a value of about $15,000$ to $20,000^\circ\text{K}$) the presence of hydrogen atoms must be considered. The proportion of these atoms under the conditions in the chromosphere may be determined by the method of Woolley & Allen (1948). Above the level where neutral hydrogen exists in appreciable proportions the pressure is found directly from the temperature and number of electrons (and hence of protons) given by the curves of figure 4. It is found that there is a marked departure from conditions of hydrostatic equilibrium; in fact, over a limited range of levels the pressure actually increases with height, indicating the presence of a supporting force greater than the gravitational force.

Recently, Giovanelli (1948) has produced a model chromosphere based on the assumption of hydrostatic equilibrium. He finds that spectroscopic observations are then explained if the temperature is $27,000^\circ\text{K}$ from 500 to about 12,000 km. This result is even more difficult to reconcile with the radio measurements than that of Redman.

It may be of interest to note that if conditions of hydrostatic equilibrium could be assumed in the chromosphere, this would provide a second relationship between N and T_e which, together with the radio equation (14), might allow a model chromosphere to be determined. The equations of hydrostatic equilibrium for a fully ionized hydrogen atmosphere are

$$dP = -gNMdy,$$

$$P = 2NkT_e,$$

where P is the pressure, M the mass of a proton and k is Boltzmann's constant. Writing $c = 2k/Mg$ and eliminating P , we have

$$\frac{c}{T_e}dT_e + \frac{c}{N}dN + dy = 0.$$

An inspection of this equation and equation (14) indicates that no solution in terms of real values of N and T_e exists.* The conclusion is that the radio observations indicate that conditions of hydrostatic equilibrium do not hold in the chromosphere.

* Thanks are due to Mr T. Pearcey for drawing attention to this fact.

The work described in this paper is part of the research programme of the Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Australia. The author wishes to thank Drs R. v. d. R. Woolley, C. W. Allen, J. L. Pawsey, D. F. Martyn and R. G. Giovanelli and Messrs K. C. Westfold and S. F. Smerd for helpful discussions.

APPENDIX 1

The expression for the absorption coefficient given by equation (1) holds only if four conditions are satisfied, namely:

(a) The dielectric constant of the medium does not depart appreciably from unity.

The dielectric constant is given by

$$\mu^2 = 1 - \frac{N\epsilon^2}{\pi m f^2},$$

so that if a value of $\mu = 0.9$ is taken as sufficiently near unity we have for a frequency of 600 Mc./sec.

$$N = 8.5 \times 10^8.$$

It is shown in §5 that the equivalent temperature of the centre of the solar disk for a frequency of 600 Mc./sec. is about 1.5×10^5 ° K, which may be taken as a reasonable estimate of the electron temperature in the region where most of the 600 Mc./sec. emission originates. Substituting these values for N and T_e in equation (2) and putting $\zeta = 0.09$, we find $\kappa = 3.5 \times 10^{-9}$. Thus a layer of thickness 3000 km. will have unit optical depth.

The above value of κ is sufficiently large to ensure complete absorption of a ray entering the solar atmosphere before μ departs greatly from unity. However, 600 Mc./sec. is considered as the lower-frequency limit of the range over which the equations derived in §5 are valid. At higher frequencies the absorption is much more rapid for a given value of μ .

(b) The collision frequency is much smaller than the wave frequency, i.e. $\nu \ll f$.

The value of ν found by combining equations (1) and (2) and putting $\nu = 0.07$ is

$$\nu = 26NT_e^{-\frac{1}{2}}.$$

The highest value of ν in this connexion occurs at a level where $T_e \sim 8000$ ° K and $N \sim 6 \times 10^9$ cm.⁻³, so that $\nu \sim 2 \times 10^5$, which is small compared to the wave frequency concerned. At higher levels ν is even smaller.

It may also be shown that nowhere do collisions of electrons with neutral atoms play an important part. For example, at the level where $N = 10^{10}$ cm.⁻³ the number of neutral hydrogen atoms per proton (as given by Saha's formula) is about 15. Since the collision cross-section for protons is some hundreds of times larger than for atoms (about 10^{-17} cm.²), the effect of the latter may be neglected.

(c) The effect of any magnetic field is negligible.

The criterion for this is $(He/mc)^2 \ll (2\pi f)^2$. Since we are not dealing with areas in the vicinity of sunspots the largest magnetic field which we need consider is of the order of 50 gauss. At a frequency of 600 Mc./sec., which is the lowest for which accuracy is claimed, the value of $\left(\frac{He}{2\pi mcf}\right)^2$ is 0.054, indicating that the effect of the magnetic field is negligible. At higher frequencies the effect is still smaller.

(d) The absorption coefficient as used assumes that the only mechanism producing absorption is that of the free-free transition of an electron when in the field of a proton. If other processes, such as those due to small changes in molecular energy, or to energy associated with nuclear spin, are present, their effects should show up as irregularities in the curve of equivalent temperature against frequency. It is unlikely, however, that within the frequency limits here considered such effects are significant. At higher frequencies absorption lines might be detected and usefully exploited. They would show up presumably as 'bright' lines in the radio spectrum, the reverse of Fraunhofer lines.

APPENDIX 2

A comparison of the theoretical distribution of solar radiation with experimental results

Experimental methods (interferometry and observations during an eclipse) are not at present capable of determining in any detail the distribution of radiation intensity over the solar disk. They provide a result which is consistent with a certain proportion of the total observed flux density originating over a uniform disk, coincident with the optical disk and the remainder originating near the limb of the optical disk. The proportion originating near the limb, expressed as a percentage P , is listed in table 4 for a number of measurements at frequencies from 600 to 24,000 Mc./sec.

The expression for the intensity of radiation over the disk given by equation (11) may be roughly checked against the experimental results by arbitrarily dividing the total radiation received into a 'disk' and a 'limb' component as follows: Divide the sun into an inner disk of radius 0.8 time that of the visible disk and an outer ring. The 'disk' component of observed flux density is equal to the radiation received from the whole solar disk at an average equivalent temperature equal to that of the inner disk. The 'limb' component is the difference between the 'disk' component and the total. Each of these quantities and hence the proportion of 'limb' component may be calculated from (11). The value of P (as percentages) are listed in table 4.

TABLE 4

observer and reference	frequency (Mc./sec.)	P (experimental)	P (from equation (11))
Piddington & Minnett (1949a)	24,000	16	9.4
Minnett & Labrum (1950)	9,400	26	14.1
Piddington & Hindman (1949)	3,000	32	20.2
Christiansen, Yabsley & Mills (1949)	600	50	30.8

The values of P found from equation (11) are all lower (by about 40 %) than the experimental values. In view of our approximate method of determining the coronal component of radiation and its distribution near the limb (neglect of radiation from the region beyond the solar limb), and also considering the rather arbitrary method of finding P from equation (11), the agreement is reasonable. Also it should be remembered that we are not primarily interested in the distribution of coronal radiation but in its total flux density.

Except at the lowest frequencies the total coronal component is only a small fraction of the observed flux density, so that errors in its determination are not important.

REFERENCES

- Allen, C. W. 1947 *Mon. Not. R. Astr. Soc.* **107**, 426.
 Baumbach, S. 1937 *Astr. Nachr.* **263**, 121.
 Chapman, S. & Cowling, T. G. 1939 *The mathematical theory of non-uniform gases*. Cambridge University Press.
 Christiansen, W. M., Yabsley, D. E. & Mills, B. Y. 1949 *Nature*, **164**, 569.
 Cillié, G. G. & Menzel, D. H. 1935 *Harv. Obs. Circ.* **410**.
 Covington, A. E. 1947 *Nature*, **159**, 405.
 Dicke, R. H. & Beringer, R. 1946 *Astrophys. J.* **103**, 375.
 Gaunt, J. A. 1930 *Phil. Trans. A*, **229**, 163.
 Ginsburg, V. L. 1946 *C.R. Acad. Sci. U.R.S.S.* **52**, 487.
 Giovanelli, R. G. 1948 *Aust. J. Sci. Res.* **1**, 360.
 Goldberg, L. 1939 *Astrophys. J.* **89**, 673.
 Henry, L. G. & Keenan, P. C. 1940 *Astrophys. J.* **91**, 625.
 Lehany, F. J. & Yabsley, D. E. 1949 *Aust. J. Sci. Res. A*, **2**, 48.
 Martyn, D. F. 1946 *Nature*, **158**, 632.
 Martyn, D. F. 1948 *Proc. Roy. Soc. A*, **193**, 44.
 Menzel, D. H. & Cillié, G. G. 1937 *Astrophys. J.* **85**, 88.
 Minnett, H. C. & Labrum, N. 1950 Solar radiation at a wavelength of 3.2 centimetres, including eclipse observations. *Aust. J. Sci. Res.* (in the Press).
 Mitchell, S. A. 1930 *Astrophys. J.* **71**, 1.
 Pawsey, J. L. & Yabsley, D. E. 1949 *Aust. J. Sci. Res. A*, **2**, 198.
 Petrie, W. J. 1944 *J. R. Astr. Soc. Can.* **38**, 137.
 Piddington, J. H. & Hindman, J. V. 1949 *Aust. J. Sci. Res. A*, **2**, no. 4.
 Piddington, J. H. & Minnett, H. C. 1949a *Aust. J. Sci. Res. A*, **2**, no. 4.
 Piddington, J. H. & Minnett, H. C. 1949b *Aust. J. Sci. Res. A*, **2**, 63.
 Redman, R. O. 1942 *Mon. Not. R. Astr. Soc.* **102**, 140.
 Sander, K. F. 1947 *Nature*, **159**, 506.
 Townes, C. H. 1947 *Astrophys. J.* **105**, 235.
 Waldmeir, M. & Muller, H. 1948 *Astr. Mitt., Zurich*, no. 155.
 Wildt, R. 1947 *Astrophys. J.* **105**, 36.
 Woolley, R. v. d. R. & Allen, C. W. 1948 *Mon. Not. R. Astr. Soc.* **108**, 292.

Entropies of mixing in certain athermal binary solutions

By E. A. GUGGENHEIM, F.R.S. AND M. L. MCGLASHAN

Department of Chemistry, University of Reading

(Received 9 March 1950—Revised 25 July 1950)

Approximate formulae for the entropy of mixing of binary athermal mixtures can be obtained by means of the model, first used by Fowler & Rushbrooke, of a lattice on which molecular species differ from one another in the number and relative orientation of the lattice points which a single molecule occupies. The previous work of Chang, Huggins, Miller and Guggenheim on open chain molecules has here been extended to molecules having the shapes of equilateral triangles, regular tetrahedra and squares.

It is found that the entropy of mixing depends less on the shapes of the molecules than on their relative sizes. Flory's formula, which corresponds formally to letting the co-ordination number z tend to infinity, gives a considerably better approximation to our values than does the formula for ideal solutions.

INTRODUCTION AND SCOPE

An ideal mixture, or ideal solution, is one for which the energy of mixing is zero, while the entropy of mixing obeys the same simple laws as for mixtures of isotopes. Real solutions may deviate from ideal solutions by different behaviour of either the energy or the entropy or both. In this paper we shall be concerned only with such mixtures as, in common with ideal mixtures, have zero energy of mixing. Such mixtures have sometimes been called 'athermal solutions'. Although we do not consider this name a particularly good one, it will be convenient to use it in preference to the name 'semi-ideal solutions' which one of us used (Guggenheim 1933) some years ago. Mixtures of chemically similar substances may be expected to form solutions which are nearly, if not quite, athermal. Examples are: (a) two different normal paraffins, (b) benzene and naphthalene, (c) benzene and diphenyl, (d) methane and neopentane. In each case the two substances are chemically similar, but have appreciably different sizes and shapes. This raises the interesting question: how is the entropy of mixing affected by the sizes and shapes of the molecules?

Until about fifteen years ago many physical chemists believed that the entropy of mixing of athermal solutions was independent of the sizes and shapes of the molecules, in other words, they believed that if such a solution was athermal, it was necessarily ideal. This view was openly challenged in a discussion held by the Faraday Society (Guggenheim 1937), at which Fowler then suggested that this view could be proved or disproved by a statistical analysis of a mixture of two kinds of molecules arranged on a lattice, each molecule of the one kind occupying two neighbouring sites of the lattice and each molecule of the other kind occupying one site. This problem was attacked by Fowler & Rushbrooke (1937), who showed that such a mixture would not be ideal. Fowler's idea of a lattice model has been used subsequently by several other authors and has proved fruitful. It is the only model which leads to explicit formulae.

As a result of researches by Chang (1939*a, b*), Miller (1942, 1943) and Guggenheim (1944, 1945) we now have formulae applicable to mixtures of any number of types

of different molecules, each type occupying a different number of lattice sites. These formulae, together with a survey of preceding work, can be found in previous papers (Guggenheim, 1944, 1945) or in a monograph by Miller (1948). While these formulae are applicable to molecules occupying any number of sites, they are restricted to molecules of such shapes as the chemist calls 'open chains' in contrast to molecules containing 'closed rings'.

It is the object of the present paper to extend the previous treatment to cover certain molecules with closed rings, namely, molecules occupying three lattice points forming an equilateral triangle, molecules occupying four lattice points forming a regular tetrahedron and molecules occupying four lattice points forming a square. The treatment will be confined to binary mixtures of one of these three types of molecules together with molecules each occupying one lattice site.

For each mixture considered, the problem can be divided sharply into two parts which may be called *configurational* and *thermodynamic* respectively. The configurational part of the problem consists in answering the following question. Suppose that a fraction θ of the lattice points is occupied by molecules each occupying r sites arranged at random. Consider a set of r sites having the same configuration as the molecules being considered. What is the ratio α of the probability that this set of r sites be occupied by one of the molecules to the probability that the r sites be all unoccupied? The thermodynamic part of the problem consists in using the formula for α to calculate the number of ways the molecules, each occupying r sites, can be arranged on the lattice. This number is, of course, the same whether the remainder of the sites are supposed empty or filled by molecules each occupying a single site. This number of possible arrangements is in turn closely related to the configurational part of the entropy of the mixture and so also to the entropy of mixing.

The first part of the paper will be devoted entirely to the configurational problems which will be completely solved before the thermodynamic aspect is given in the second part of the paper. It seems just possible that some mathematician, not interested in thermodynamics, may find the treatment of the configurational problems worthy of attention, criticism and improvement.

PART I. CONFIGURATIONAL

1. METHOD

The method which will be used is essentially that used in a previous paper (Guggenheim 1945) which is a simplified version of the method previously used by Chang (1939*a, b*) and by Miller (1942, 1943). We shall, however, be able to introduce a further simplification, which will be described below.

2. TERMINOLOGY AND NOTATION

It is convenient to call molecules each occupying one, two, three, four, r sites respectively monomers, dimers, trimers, tetramers, r -mers. We shall speak of an r -mer as consisting of r elements each occupying one site. We denote by θ the fraction

of sites occupied by the r -mers ($r \neq 1$), so that $1 - \theta$ denotes the fraction of sites which at this stage we regard as vacant, in part II these sites will be considered as occupied by monomers. We denote by z the number of sites which are immediate neighbours of any given site. If then one element of an r -mer has been placed on a site there are z ways in which a second element, adjoining the one already placed, can be placed on the lattice. When these two elements have been placed, the molecule as a whole can still be placed in a number of ways which we denote by z' . We denote by ρ the number of ways in which the molecule as a whole can be placed after one of its elements has been placed. ρ is equal to the product of z and z' divided by a small integer depending on the shape of the molecule; this divisor is 1 for a straight molecule, 2 for an equilateral triangle or square molecule and 3 for a tetrahedral molecule. Values of z and z' will be stated when required.

We shall also use a quantity κ , related to ρ and always smaller than ρ , defined as follows. Consider a group of sites congruent with the r -mer and let one element of an r -mer be placed on one site of the group. Then κ denotes the number of ways the r -mer as a whole can be placed without using any other sites of the given group. Take as a simple example equilateral triangular molecules on a plane triangular lattice: $z = 6$, $z' = 2$, $\rho = \frac{1}{2}zz' = 6$, and $\kappa = 3$.

We shall denote by α the ratio of the probability that a group of sites, congruent with the r -mer, be wholly occupied by one molecule to the probability that the group be completely vacant.

Our problem then is, for molecules of each shape considered, to find a relation between α and θ .

3. DIMERS

It is expedient to begin by considering the problem of placing dimers on a lattice, since this easy example will serve to illustrate the general method and to show up its greater power and simplicity than the earlier versions used by Chang (1939*a, b*) and by Miller (1942, 1943).

TABLE 1. DIMERS

reference number	configuration		relative probability
	a	b	
1			α
2	X	X	ϵ^2
3	X	0	ϵ
3'	0	X	ϵ
4	0	0	1

We accordingly consider a pair of sites ab . Table 1 contains a list of the several possible manners of their occupation with their relative probabilities. The first column merely gives a number by which to refer to each type of configuration. The second column specifies the configuration or type of configuration. The configuration 1, denoted by a line joining the two sites, is the one in which a molecule occupies the pair of sites. The symbol X denotes that a site is occupied by one element of a molecule, while its other element is on some site other than a, b . The symbol 0 denotes that a site is vacant. The third column gives ratios of probabilities for the several configurations

or types of configuration, the value for configuration 4 with both sites empty being unity by convention. The two configurations 3 and 3' clearly have equal relative probabilities which are denoted by ϵ . The essential assumption or approximation of the treatment is that, if a and b are not occupied by the same molecule, then the relative probabilities of being occupied or vacant are independent for the two sites. This leads to a relative probability ϵ^2 for configuration 2.

We note that each factor ϵ represents $z-1$ orientations of a molecule. In other words, a single orientation of a molecule occupying a or b , but not both, is represented by a factor $\epsilon/(z-1)$. The probability, represented by α , that a molecule occupies the pair of sites ab must be equal to the probability that a molecule occupies the site a and some other specified neighbouring site. This condition is expressed by the equation

$$\alpha = \frac{\epsilon}{z-1} (\epsilon + 1). \quad (3.1)$$

Equations obtained by this type of reasoning will be called equivalence relations.

As we have already mentioned, $\epsilon/(z-1)$ is the relative probability of occupation of a by a molecule having a specified orientation. Consequently $z\epsilon/(z-1)$ is equal to the relative probability of occupation of a regardless of orientation. In other words, the probability that the site a , or any other chosen site, is occupied is $z\epsilon/(z-1)$ times the probability of its being empty. But these two probabilities are evidently to one another in the ratio θ to $1-\theta$. Hence we have

$$\frac{z}{z-1} \epsilon = \frac{\theta}{1-\theta}. \quad (3.2)$$

Eliminating ϵ between (3.1) and (3.2) we obtain

$$\alpha = \frac{1}{z} \frac{\theta}{1-\theta} \frac{1-\theta/z}{1-\theta}, \quad (3.3)$$

which is the required relation between α and θ .

Formula (3.3) is nowise new, having been obtained by Chang (1939*a, b*) and by Miller (1942), but the derivation given here is much shorter and more direct than theirs. There are four contributory causes to the greater simplicity of the new derivation. In the first place Chang and Miller began by considering the more general and much more complicated problem in which the several configurations had different energies; they then treated the case of an athermal system, in which all configurations have equal energies, as a particular example. In the second place Chang considered a group of altogether $2z$ sites and Miller a group of $z+1$ sites, whereas here we are considering only two sites. In the third place both Chang and Miller distinguished between central sites and neighbouring sites, whereas here the two sites are treated on a par. In the fourth place if we had copied the technique used by Chang and Miller our derivation of formula (3.2) would have been as follows. We equate $\theta/(1-\theta)$ to the ratio of the sum of all relative probabilities for configurations in which a is occupied to the sum of those for configurations in which a is empty. This gives us

$$\frac{\theta}{1-\theta} = \frac{\alpha + \epsilon^2 + \epsilon}{\epsilon + 1}. \quad (3.4)$$

When we substitute from (3.1) into (3.4) we recover (3.2) by an unnecessarily circuitous route. This alternative derivation admittedly affords a useful check on the correctness and mutual consistency of (3.1) and (3.2).

These simplifications may at first sight seem unimportant, but it is in fact their introduction which has made tractable the more complicated examples which we are now about to study. Before proceeding to these, it is convenient to introduce an abbreviation for use in tables of configurations.

If we use the symbol U to denote either X or 0 , then in place of table 1 we have the abbreviated table 1 A.


TABLE 1 A. DIMERS

reference number	configuration	relative probability
1	$\frac{a}{U} \frac{b}{U}$	α
2	$U \quad U$	$(\epsilon + 1)^2$

4. EQUILATERAL TRIANGLES

We are now ready to attack the more interesting problem of trimers each occupying three sites forming an equilateral triangle. The same treatment applies either to a plane triangular lattice ($z = 6$, $z' = 2$) or to a spatial close-packed lattice ($z = 12$, $z' = 4$). We consider a group of three sites a , b , c forming an equilateral triangle. Its several possible manners of occupation, together with their relative probabilities, are summarized in table 2. We use the abbreviated notation of table 1 A, where

TABLE 2. TRIANGULAR TRIMERS

reference number	configuration	relative probability
1		α
2	$\frac{U}{U}$	$3(z' - 1) \zeta(\epsilon + 1)$
3	$\frac{U}{U} \quad \frac{U}{U}$	$(\epsilon + 1)^3$

U denotes either X (occupied) or 0 (vacant). We continue to use the approximation of assuming that if two sites, say a and b , are not occupied by the same molecule then the relative probabilities of being occupied or vacant are independent for the two sites. The fact that this is only an approximation is more obvious in the present case than in that of dimers. Configurations 1 and 3 call for no further comment. In the set of configurations referred to as 2, two, and only two, of the three sites a , b , c are occupied by the same molecule. In the formula for the relative probability the factor 3 represents the three possible ways of choosing a pair of sites out of the three, the

factor $z' - 1$ represents the number of possible configurations of a molecule specified as occupying two sites say a, b but not the third c . The factor $\epsilon + 1$ takes care of the third site. If, then, we completely specify the configuration of the molecule occupying two of the sites, leaving unspecified whether the third site is empty or occupied, the relative probability is $\zeta(\epsilon + 1)$. This may be regarded as the definition of ζ . The parameter ϵ represents the occupation by a molecule of one site, and one only, of the group of three and comprises a number κ of alternative configurations of this molecule. Hence any one of these alternative configurations is represented by ϵ/κ . Evidently each completely specified configuration of a molecule has the same probability whether this molecule occupies all three sites of the group or two of them or only one. This gives the equivalence relations

$$\alpha = \zeta(\epsilon + 1) = \frac{\epsilon}{\kappa} \{ (z' - 1) \zeta + (\epsilon + 1)^2 \}. \quad (4.1)$$

If we solve for ζ and α in terms of ϵ we obtain

$$\zeta = \frac{\epsilon}{\frac{\kappa \kappa - z' + 1}{\kappa} \epsilon + 1} \frac{(\epsilon + 1)^2}{\epsilon + 1}, \quad (4.2)$$

$$\alpha = \frac{\epsilon}{\frac{\kappa \kappa - z' + 1}{\kappa} \epsilon + 1} \frac{(\epsilon + 1)^3}{\epsilon + 1}. \quad (4.3)$$

Since, as already mentioned, ϵ/κ is the relative probability of occupation of a given site by a molecule having a specified orientation, it follows that the total relative probability of occupation of a given site is $\rho\epsilon/\kappa$. Since this is true for any and every site we must have

$$\frac{\rho}{\kappa} \epsilon = \frac{\theta}{1 - \theta}. \quad (4.4)$$

Substituting (4.4) into (4.3) we obtain

$$\alpha = \frac{\theta}{\rho} \frac{\left(1 - \frac{\rho - \kappa}{\rho} \theta\right)^3}{(1 - \theta)^3 \left(1 - \frac{\rho - \kappa + z' - 1}{\rho} \theta\right)}. \quad (4.5)$$

According to the definitions of ρ and κ their values for equilateral triangular molecules are

$$\rho = \frac{1}{2} z z', \quad \kappa = \rho - 2 z' + 1. \quad (4.6)$$

Using (4.6) in (4.5) we obtain finally

$$\alpha = \frac{\theta}{\frac{1}{2} z z'} \frac{\left(1 - \frac{2 z' - 1}{\frac{1}{2} z z'} \theta\right)^3}{(1 - \theta)^3 \left(1 - \frac{3 z' - 2}{\frac{1}{2} z z'} \theta\right)}, \quad (4.7)$$

valid either for a planar triangular lattice with $z = 6$, $z' = 2$ or for a spatial close packed lattice with $z = 12$, $z' = 4$.

Incidentally, from the relative probabilities taken from table 2 we can write down the chance that a single site be empty, that two neighbouring sites be empty or that three sites forming an equilateral triangle be empty. If we do this and use the equivalence relations (4.1) and the relation (4.4) between ϵ and θ , we eventually find for the probabilities:

One site empty: $1 - \theta.$ (4.8)

Pair of sites both empty:
$$\frac{(1 - \theta)^2 \left(1 - \frac{3z' - 2}{\frac{1}{2}zz'} \theta\right)}{\left(1 - \frac{2z' - 1}{\frac{1}{2}zz'} \theta\right)^2}.$$
 (4.9)






Triangle of sites all empty:
$$\frac{(1 - \theta)^3 \left(1 - \frac{3z' - 2}{\frac{1}{2}zz'} \theta\right)}{\left(1 - \frac{2z' - 1}{\frac{1}{2}zz'} \theta\right)^3}.$$
 (4.10)

The result given by (4.8) is physically obvious, but this method of deriving (4.8) forms a valuable check on the self-consistency of the formulae. By contrast we have not found any more direct derivation of either (4.9) or (4.10).

5. REGULAR TETRAHEDRA

We now pass on to the problem of placing regular tetrahedral tetramers on a close packed lattice, for which $z = 12$, $z' = 2$.

TABLE 3. TETRAHEDRAL TETRAMERS

reference number	configuration	relative probability
1		α
2		$3(z' - 1)^2 \xi = 3\xi$
3		$6(z' - 1) \xi(\epsilon + 1)^2 = 6\xi(\epsilon + 1)^2$
4		$(\epsilon + 1)^4$
		

We consider a group of four sites, $abcd$ forming a regular tetrahedron. Its several possible manners of occupation, together with their relative probabilities, are summarized in table 3. The relative position of the four sites a, b, c, d can be shown only schematically; the pairs ac and bd are in all respects equivalent to the pairs ab, bc, cd and da . As previously the symbol U denotes either occupied by

a molecule that does not occupy any other sites of the group or vacant; these two alternatives have probabilities in the ratio ϵ to 1. The first and fourth configurations call for no comment. In the second set of configurations two molecules occupy opposite edges of the tetrahedron. The factor 3 in the relative probability represents the three pairs of opposite sides. Each of the factors $z' - 1 = 1$ represents the number of ways a molecule can occupy a given pair of sites, say ab , without occupying either of the other two c, d . Then ξ is by definition the probability that a molecule occupies a given pair of sites, say ab , with a specified orientation and simultaneously another molecule occupies the other pair of sites cd with a specified orientation. In the third set of configurations one molecule occupies two sites of the group while the other two sites may be occupied or empty, but are not occupied by the same molecule. The factor 6 corresponds to the six edges of the tetrahedron. The factor $z' - 1 = 1$ represents the number of ways a molecule can occupy two of the sites, say a, b , without occupying either of the other two, c, d . The factor ζ is defined as the relative probability of a molecule occupying two given sites, say a, b , with a specified orientation when the other two sites c, d are empty. The two factors $(\epsilon + 1)$ take care of the other two sites.

By considerations closely similar to those applying to triangular molecules we obtain the equivalence relations

$$\alpha = \xi + \zeta(\epsilon + 1)^2 = \frac{\epsilon}{\kappa} \{3\zeta(\epsilon + 1) + (\epsilon + 1)^3\}. \quad (5.1)$$

There is also another independent equivalence relation

$$\xi = \frac{\epsilon^2}{\kappa^2} \{\zeta + (\epsilon + 1)^2\}, \quad (5.2)$$

expressing the fact that by using only two of the four sites $abcd$ one can obtain configurations geometrically equivalent to those of the second set. When we solve (5.1) and (5.2) for ζ, ξ, α in terms of ϵ we obtain

$$\zeta = \frac{\epsilon}{\kappa} \frac{(\epsilon + 1)^2 \left(\epsilon + 1 - \frac{\epsilon}{\kappa} \right)}{(\epsilon + 1)^2 - 3 \frac{\epsilon}{\kappa} (\epsilon + 1) + \frac{\epsilon^2}{\kappa^2}}, \quad (5.3)$$

$$\xi = \frac{\epsilon^2}{\kappa^2} \frac{(\epsilon + 1)^3 \left(\epsilon + 1 - 2 \frac{\epsilon}{\kappa} \right)}{(\epsilon + 1)^2 - 3 \frac{\epsilon}{\kappa} (\epsilon + 1) + \frac{\epsilon^2}{\kappa^2}}, \quad (5.4)$$

$$\alpha = \frac{\epsilon}{\kappa} \frac{(\epsilon + 1)^3 \left\{ (\epsilon + 1)^2 - 2 \frac{\epsilon^2}{\kappa^2} \right\}}{(\epsilon + 1)^2 - 3 \frac{\epsilon}{\kappa} (\epsilon + 1) + \frac{\epsilon^2}{\kappa^2}}. \quad (5.5)$$

We have the usual relation between ϵ and θ

$$\epsilon = \frac{\kappa}{\rho} \frac{\theta}{1 - \theta} \quad (5.6)$$

with

$$\rho = \frac{1}{3} z z', \quad \kappa = \rho - 3 z' + 2. \quad (5.7)$$

Substituting (5.6) and (5.7) into (5.5), we obtain

$$\alpha = \frac{\theta \left(1 - \frac{3z' - 2}{\rho} \theta\right)^3 \left(1 - \frac{6z' - 4}{\rho} \theta + \frac{9z'^2 - 12z' + 2}{\rho^2} \theta^2\right)}{(1 - \theta)^4 \left(1 - \frac{6z' - 1}{\rho} \theta + \frac{9z'^2 - 3z' - 1}{\rho^2} \theta^2\right)}. \quad (5.8)$$

When we put $z = 12$, $z' = 2$, $\rho = 8$ corresponding to a close packed lattice, we obtain





$$\alpha = \frac{\theta (1 - \frac{1}{2}\theta)^3 (1 - \theta + \frac{7}{32}\theta^2)}{8(1 - \theta)^4 (1 - \frac{11}{8}\theta + \frac{29}{64}\theta^2)}. \quad (5.9)$$

If we take from table 3 the ratio of the sum of all the relative probabilities when the site a is occupied to the sum of all those when the site a is vacant, use the equivalence relations (5.1), (5.2) and the relation (5.6) between ϵ and θ we obtain eventually just $\theta/(1 - \theta)$ thus checking the self-consistency of our formulae. We can by a similar procedure obtain formulae in terms of θ for the chance that a pair of sites, a triangle of sites or a tetrahedron of sites be vacant, but we shall not quote the results.

6. SQUARES

As our last example we shall consider square molecules occupying four sites on a simple cubic lattice with $z = 6$, $z' = 4$. We consider a group of four sites $abcd$ forming a square. Its several manners of occupation, together with their relative

TABLE 4. SQUARE TETRAMERS

reference number	configuration	relative probability
1		α
2		$2(z' - 1)^2 \xi$
3		$4(z' - 1) \zeta(\epsilon + 1)^2$
4		$(\epsilon + 1)^4$

probabilities, are summarized in table 4. The first and fourth configurations call for no comment. In the second set of configurations each of two molecules occupy two opposite sides of the square. In the relative probability the factor 2 represents the two pairs of opposite sides of the square and the two factors $z' - 1$ the alternative orientations of each of the two molecules. In the third set of configurations one, and

only one, side of the square is occupied by the same molecule. In the relative probability the factor 4 represents the four alternative sides of the square and the factor $z' - 1$ the alternative orientations of the molecule.

The equivalence relations are

$$\alpha = (z' - 1)\xi + \zeta(\epsilon + 1)^2 = \frac{\epsilon}{\kappa}(\epsilon + 1)\{2(z' - 1)\xi + (\epsilon + 1)^2\}, \quad (6.1)$$

$$\xi = \frac{\epsilon^2}{\kappa^2}\{(z' - 1)\xi + (\epsilon + 1)^2\}. \quad (6.2)$$

Solving for ζ , ξ , α in terms of ϵ we obtain

$$\zeta = \frac{\epsilon}{\kappa} \frac{(\epsilon + 1)^2}{\frac{\kappa - z' + 1}{\kappa} \epsilon + 1}, \quad (6.3)$$

$$\xi = \frac{\epsilon^2}{\kappa^2} \frac{(\epsilon + 1)^3}{\frac{\kappa - z' + 1}{\kappa} \epsilon + 1}, \quad (6.4)$$

$$\alpha = \frac{\epsilon}{\kappa} \frac{(\epsilon + 1)^3 \left(\frac{\kappa + z' - 1}{\kappa} \epsilon + 1 \right)}{\frac{\kappa - z' + 1}{\kappa} \epsilon + 1}. \quad (6.5)$$

We have the usual relation between ϵ and θ

$$\epsilon = \frac{\kappa}{\rho} \frac{\theta}{1 - \theta}. \quad (6.6)$$

Substituting (6.6) into (6.5) we obtain

$$\alpha = \frac{\theta}{\rho} \frac{\left(1 - \frac{\rho - \kappa}{\rho} \theta\right)^3 \left(1 - \frac{\rho - \kappa - z' + 1}{\rho} \theta\right)}{(1 - \theta)^4 \left(1 - \frac{\rho - \kappa + z' - 1}{\rho} \theta\right)}. \quad (6.7)$$

For square tetramers we have

$$\rho = \frac{1}{2}zz', \quad \kappa = \rho - 2z' + 1, \quad (6.8)$$

so that (6.7) reduces to

$$\alpha = \frac{\theta}{\frac{1}{2}zz'} \frac{\left(1 - \frac{2z' - 1}{\frac{1}{2}zz'} \theta\right)^3 \left(1 - \frac{2}{z} \theta\right)}{(1 - \theta)^4 \left(1 - \frac{3z' - 2}{\frac{1}{2}zz'} \theta\right)}, \quad (6.9)$$

wherein we have to use the values $z = 6$, $z' = 4$ for a simple cubic lattice.

7. NATURE OF APPROXIMATION

As already mentioned the essential approximation in our treatment is the assumption that when two sites are not occupied by the same molecule the probabilities of being occupied or vacant are independent for the two sites. It is easy to see by an example that this assumption is at least sometimes false. Consider the triangular

group of sites abc used in §4. Let d denote another site forming an equilateral triangle with b, c . Then when we state that a molecule occupies the site b , but not a or c , it may also occupy the site d . Likewise, when we state that a molecule occupies the site c , but not a or b , it may also occupy the site d . When we state that two different molecules occupy the sites b and c , without occupying a , then either of them, but *not both*, may also occupy the site d . Hence the manners of occupation of sites b and c are not independent even when they are not occupied by the same molecule.

It is not easy to estimate the error due to this approximation. Its seriousness will almost certainly vary from one case to another. The best method of estimating the error is probably to modify the whole treatment in such a manner that the approximation is replaced by a less inaccurate one. This is in principle possible, but the formulae obtained are likely to be of greatly increased complexity. The equivalence relations will probably not be soluble except numerically. We are planning to study the case of dimers and possibly triangles on a triangular lattice to a better approximation and to compare the results with those of the present approximation.

The formulae obtained, though admittedly not accurate, are at least self-consistent in the cases studied above. They, however, break down in a curiously surprising manner when applied to square molecules on a planar square lattice. We derived formulae (6.5) and (6.7) for square molecules without specifying particular values of z or z' . We might therefore expect these formulae to be applicable as well to a planar square lattice as to a simple cubic lattice. For a planar square lattice we have

$$z = 4, \quad z' = 2, \quad \rho = \frac{1}{2}zz' = 4, \quad \kappa = \frac{1}{2}zz' - 2z' + 1 = 1.$$

When we substitute these values into (6.5) we find that the coefficient of ϵ in the last factor of the denominator becomes zero, so that this factor reduces to unity. Although this occurrence rather strikingly alters the dependence of α on θ , it may at first sight seem believable. If, however, we examine the behaviour at saturation we shall find that the conclusions are physically unacceptable.

Let us then examine our formulae when θ tends to 1. Evidently ϵ tends to infinity. The ratios $\alpha : \xi : \zeta\epsilon^2 : \epsilon^4$ have finite values, as physically they must have, except when $k - z' + 1$ vanishes. For squares on a planar square lattice $k - z' + 1$ vanishes and then $\alpha : \epsilon^5$ becomes finite, so that $\alpha : \epsilon^4$ becomes infinite. This would mean that when the lattice is fully occupied, the four sites a, b, c, d can never be occupied by four different molecules. Such a conclusion is physically absurd. Thus in the particular case of square molecules on a planar square lattice our formulae are physically unacceptable. We have not been able to find any explanation why this should happen in this one case but in none of the others examined.

PART II. THERMODYNAMICS

8. NOTATION

In addition to the symbols defined in part I, the following symbols will also be used:

N_s	total number of lattice sites
$N_1 = (1 - \theta)N_s$	number of monomers

$N_r = \theta N_s / r$	number of r -mers
$x = N_r / (N_1 + N_r)$	mole fraction of r -mer
k	Boltzmann's constant
R	gas constant
T	absolute temperature
F	Helmholtz function (free energy)
G	Gibbs function
S	entropy
μ_1	chemical potential of monomer (per molecule)
μ_r	chemical potential of r -mer (per molecule)
λ_1	absolute activity of monomer
λ_r	absolute activity of r -mer
f_1	partition function of monomer occupying a given site
f_r	partition function of r -mer occupying a given set of sites
σ	symmetry number of r -mer
$\phi_r = f_r \sigma$	

9. GENERAL RELATIONS

We must now consider how to obtain formulae for the entropy of mixing in terms of the quantity which has been denoted by α . This quantity is closely related to the thermodynamic quantities λ called the *absolute activities* (Fowler & Guggenheim 1939; Guggenheim 1949), whose properties we shall briefly recall. We begin with one of Gibbs's fundamental relations, which for a binary mixture of a monomer and an r -mer take the form

$$dG = -SdT + VdP + \mu_1 dN_1 + \mu_r dN_r. \quad (9.1)$$

Since we are considering only condensed (non-gaseous) phases, the distinction between G and F is trivial; we accordingly replace G by the somewhat more familiar F . As we shall be concerned only with variations of compositions at given temperature and pressure, we omit the terms in dT and dP , so that (9.1) reduces to

$$dF = \mu_1 dN_1 + \mu_r dN_r. \quad (9.2)$$

The absolute activities λ_1 , λ_r are related to the respective chemical potentials μ_1 , μ_r through the definitions

$$\mu_1 = kT \log \lambda_1, \quad \mu_r = kT \log \lambda_r. \quad (9.3)$$

We may accordingly replace (9.2) by

$$\frac{1}{kT} dF = \log \lambda_1 dN_1 + \log \lambda_r dN_r. \quad (9.4)$$

Changing from the variables N_1 , N_r to the variables N_s , θ we have

$$\frac{1}{N_s kT} dF = \left\{ (1-\theta) \log \lambda_1 + \frac{\theta}{r} \log \lambda_r \right\} d \log N_s + \left\{ -\log \lambda_1 + \frac{1}{r} \log \lambda_r \right\} d\theta, \quad (9.5)$$

or, for constant N_s ,

$$\frac{1}{N_s kT} dF = \frac{1}{r} \log \frac{\lambda_r}{\lambda_1^r} d\theta. \quad (9.6)$$

We shall now show that formula (9.6) leads directly to a relation between the entropy of mixing and α .

Either by use of grand partition functions (Miller 1948) or by elementary considerations based on the principle of detailed balancing (Guggenheim 1944) it can be shown that α is related to λ_1 , λ_r by

$$\alpha = \frac{\lambda_r f_r}{(\lambda_1 f_1)^r} = \frac{\lambda_r \phi_r}{\sigma (\lambda_1 f_1)^r}. \quad (9.7)$$

The partition functions f_1, f_r take account of the vibrational motions of the centres of mass of the molecules, while f_r contains further factors for the librations and internal vibrations of the r -mer molecules. In order to emphasize that f_r always contains a factor $1/\sigma$ it is useful to write f_r as ϕ_r/σ .

When we substitute (9.7) into (9.6) we obtain the simple relation between F and α

$$\frac{1}{N_s kT} dF = \frac{1}{r} \log \frac{\alpha \sigma f_1^r}{\phi_r} d\theta. \quad (9.8)$$

To obtain F we have only to substitute into (9.8) the relevant formula for α obtained in part I and integrate from 0 to θ . Each of the formulae for α contains a factor ρ , so that the quantity whose logarithm occurs in (9.8) contains a factor $(\rho/\sigma)^{-1}$. This is a convenient place to mention that the quantity now represented by ρ/σ was in previous papers represented by ρ . The notation used previously was chosen to facilitate comparison with the earlier work of Chang, but the present notation seems more reasonable. In any case this factor disappears from the formulae for the free energy of mixing and for the entropy of mixing

If we regard the free energy as a function of N_s and θ , then the free energy of mixing ΔF is defined by

$$\Delta F = F(N_s, \theta) - (1 - \theta) F(N_s, 0) - \theta F(N_s, 1). \quad (9.9)$$

Comparing (9.8) with (9.9) we see that ΔF takes the form

$$\frac{\Delta F}{N_s kT} = \frac{1}{r} \left\{ \int_0^\theta \log \alpha d\theta - \theta \int_0^1 \log \alpha d\theta \right\}. \quad (9.10)$$

Since we are concerned only with athermal solutions, having by definition zero energies of mixing, the entropy of mixing ΔS is given by

$$\frac{\Delta S}{N_s k} = \frac{1}{r} \left\{ - \int_0^\theta \log \alpha d\theta + \theta \int_0^1 \log \alpha d\theta \right\}. \quad (9.11)$$

Formula (9.11) gives the entropy of mixing of a mixture containing $N_1 = (1 - \theta) N_s$ molecules of monomer and $N_r = \theta N_s / r$ molecules of r -mer. To obtain the molar entropy of mixing ΔS_m we have to divide ΔS by the total number of molecules, namely, $(1 - \theta + \theta/r) N_s$, and multiply by Avogadro's number which is equal to R/k . We thus obtain

$$\frac{\Delta S_m}{R} = \frac{1}{r - (r-1)\theta} \left\{ - \int_0^\theta \log \alpha d\theta + \theta \int_0^1 \log \alpha d\theta \right\}. \quad (9.12)$$

In the following section formula (9.12) will be used to calculate molar entropies of mixing from the formulae for α derived in part I.

10. FORMULAE FOR ENTROPIES OF MIXING

We now apply formula (9.12) in turn to the several kinds of mixtures considered in part I.

Dimers. We write formula (3.3) in the form

$$\log z = \log \theta + \log(z - \theta) - 2 \log(1 - \theta) - 2 \log z. \quad (10.1)$$

Substituting this into (9.12) and performing the integrations we obtain

$$\frac{\Delta S_m}{R} = (2 - \theta)^{-1} \{ -\theta \log \theta - 2(1 - \theta) \log(1 - \theta) + (z - \theta) \log(z - \theta) - (1 - \theta) z \log z - \theta(z - 1) \log(z - 1) \}, \quad (10.2)$$

in agreement with Chang (1939*a*, *b*).

For a cubic lattice ($z = 6$) we have

$$\frac{\Delta S_m}{R} = (2 - \theta)^{-1} \{ -\theta \log \theta - 2(1 - \theta) \log(1 - \theta) + (6 - \theta) \log(6 - \theta) - (1 - \theta) 6 \log 6 - \theta 5 \log 5 \}. \quad (10.3)$$

Equilateral triangles. Substituting (4.7) into (9.12) and performing the integrations we obtain

$$\begin{aligned} \frac{\Delta S_m}{R} = (3 - 2\theta)^{-1} & \left\{ -\theta \log \theta - 3(1 - \theta) \log(1 - \theta) + 3 \left(\frac{\rho}{2z' - 1} - \theta \right) \log \left(\frac{\rho}{2z' - 1} - \theta \right) \right. \\ & - \left(\frac{\rho}{3z' - 2} - \theta \right) \log \left(\frac{\rho}{3z' - 2} - \theta \right) - (1 - \theta) \left[\frac{3\rho}{2z' - 1} \log \frac{\rho}{2z' - 1} - \frac{\rho}{3z' - 2} \log \frac{\rho}{3z' - 2} \right] \\ & \left. - \theta \left[3 \left(\frac{\rho}{2z' - 1} - 1 \right) \log \left(\frac{\rho}{2z' - 1} - 1 \right) - \left(\frac{\rho}{3z' - 2} - 1 \right) \log \left(\frac{\rho}{3z' - 2} - 1 \right) \right] \right\}, \quad (10.4) \end{aligned}$$

where $\rho = \frac{1}{2}zz'$. On a close-packed space lattice $z = 12$, $z' = 4$, $\rho = 24$. When we put these values into (10.4) we obtain

$$\begin{aligned} \frac{\Delta S_m}{R} = (3 - 2\theta)^{-1} & \{ -\theta \log \theta - 3(1 - \theta) \log(1 - \theta) + 3 \left(\frac{24}{7} - \theta \right) \log \left(\frac{24}{7} - \theta \right) \\ & - \left(\frac{12}{5} - \theta \right) \log \left(\frac{12}{5} - \theta \right) - (1 - \theta) \left(\frac{72}{7} \log \frac{24}{7} - \frac{12}{5} \log \frac{12}{5} \right) - \theta \left(\frac{51}{7} \log \frac{17}{7} - \frac{7}{5} \log \frac{7}{5} \right) \}. \end{aligned} \quad (10.5)$$

Tetrahedra. As the formulae expressed in terms of z , z' are cumbersome we shall proceed directly to those in which the numerical values $z = 12$, $z' = 2$ have already been inserted. Formula (5.9) is conveniently rewritten as

$$\alpha = \frac{\theta^{(1 - \frac{1}{2}\theta)^3} \left(1 - \frac{\theta}{a} \right) \left(1 - \frac{\theta}{b} \right)}{(1 - \theta)^4 \left(1 - \frac{\theta}{c} \right) \left(1 - \frac{\theta}{d} \right)}, \quad (10.6)$$

where $a = \frac{8}{4 - \sqrt{2}}$, $b = \frac{8}{4 + \sqrt{2}}$, $c = \frac{16}{11 - \sqrt{5}}$, $d = \frac{16}{11 + \sqrt{5}}$.

Substituting (10.6) into (9.12) and performing the integrations, we obtain

$$\begin{aligned} \frac{\Delta S_m}{R} = & (4-3\theta)^{-1} \{ -\theta \log \theta - 4(1-\theta) \log (1-\theta) + 3(2-\theta) \log (2-\theta) \\ & + (a-\theta) \log (a-\theta) + (b-\theta) \log (b-\theta) - (c-\theta) \log (c-\theta) - (d-\theta) \log (d-\theta) \\ & - (1-\theta) [6 \log 2 + a \log a + b \log b - c \log c - d \log d] \\ & - \theta [(a-1) \log (a-1) + (b-1) \log (b-1) - (c-1) \log (c-1) - (d-1) \log (d-1)] \}. \end{aligned} \quad (10.7)$$

Squares. Substituting (6.9) into (9.12) and integrating we obtain

$$\begin{aligned} \frac{\Delta S_m}{R} = & (4-3\theta)^{-1} \left\{ -\theta \log \theta - 4(1-\theta) \log (1-\theta) + 3 \left(\frac{\rho}{2z'-1} - \theta \right) \log \left(\frac{\rho}{2z'-1} - \theta \right) \right. \\ & + \left(\frac{\rho}{z'} - \theta \right) \log \left(\frac{\rho}{z'} - \theta \right) - \left(\frac{\rho}{3z'-2} - \theta \right) \log \left(\frac{\rho}{3z'-2} - \theta \right) \\ & - (1-\theta) \left[3 \frac{\rho}{2z'-1} \log \frac{\rho}{2z'-1} + \frac{\rho}{z'} \log \frac{\rho}{z'} - \frac{\rho}{3z'-2} \log \frac{\rho}{3z'-2} \right] \\ & - \theta \left[3 \left(\frac{\rho}{2z'-1} - 1 \right) \log \left(\frac{\rho}{2z'-1} - 1 \right) \right. \\ & \left. \left. + \left(\frac{\rho}{z'} - 1 \right) \log \left(\frac{\rho}{z'} - 1 \right) - \left(\frac{\rho}{3z'-2} - 1 \right) \log \left(\frac{\rho}{3z'-2} - 1 \right) \right] \right\}, \end{aligned} \quad (10.8)$$

where $\rho = \frac{1}{2}zz'$. For a simple cubic lattice $z = 6, z' = 4, \rho = 12$. Inserting these values into (10.8) we obtain

$$\begin{aligned} \frac{\Delta S_m}{R} = & (4-3\theta)^{-1} \{ -\theta \log \theta - 4(1-\theta) \log (1-\theta) + 3 \left(\frac{12}{7} - \theta \right) \log \left(\frac{12}{7} - \theta \right) \\ & + (3-\theta) \log (3-\theta) - \left(\frac{6}{5} - \theta \right) \log \left(\frac{6}{5} - \theta \right) - (1-\theta) \left(\frac{36}{7} \log \frac{12}{7} + 3 \log 3 - \frac{6}{5} \log \frac{6}{5} \right) \\ & - \theta \left(\frac{12}{7} \log \frac{12}{7} + 2 \log 2 - \frac{1}{5} \log \frac{1}{5} \right) \}. \end{aligned} \quad (10.9)$$

11. EARLIER WORK

In order to discuss adequately the few previous attacks on the problems of the present paper it is necessary first to say a few words about the similar problem for open-chain molecules. For such molecules the relation between α and θ can be simply expressed by introducing a number q defined as follows.

Consider an r -mer occupying a given group of r sites. Each of these r sites has z neighbouring sites, some occupied by other elements of the same molecule. We denote by zq the number of pairs of neighbouring sites of which one is a member of the group occupied by the given molecule and the other is not. Then for an open chain molecule q is related to r by

$$z(r-q) = 2(r-1). \quad (11.1)$$

For molecules, such as equilateral triangles and squares, with a single closed ring, the relation is

$$z(r-q) = 2r. \quad (11.2)$$

For tetrahedra the relation can be written

$$z(r-q) = 3r. \quad (11.3)$$

For open chain r -mers the relation between α and θ is

$$\alpha = \frac{\sigma\theta \left(1 - \frac{r-q}{r}\theta\right)^{r-1}}{\rho r(1-\theta)^r}, \quad (11.4)$$

which leads to the formula for the molar entropy of mixing

$$\begin{aligned} \frac{\Delta S_m}{R} = \left(1 - \frac{r-1}{r}\theta\right)^{-1} & \left\{ -\frac{\theta}{r} \log \theta - (1-\theta) \log (1-\theta) \right. \\ & \left. + \frac{r-1}{r} \left(\frac{r}{r-q} - \theta\right) \log \left(1 - \frac{r-q}{r}\theta\right) - \theta \frac{q(r-1)}{r(r-q)} \log \frac{q}{r} \right\}. \end{aligned} \quad (11.5)$$

Formula (11.5) was first derived by Huggins (1942). Actually Huggins's formula contains a complicated small quantity denoted by f_0 ; if one sets $f_0 = 0$ in Huggins's formula it reduces to (11.5). Formula (11.5) was also tentatively proposed by Miller (1943) as a generalization of formulae already obtained for the special cases of dimers (Chang 1939*a, b*) and straight trimers (Miller 1942). A simple derivation of (11.4) was subsequently given by one of us (Guggenheim 1944).

A less accurate, but simpler, formula was proposed independently by Flory (1942). It can be obtained formally from Huggins's formula by making $z \rightarrow \infty$. This means that in (11.5) we make $(r-q)/r \rightarrow 0$. The resulting formula is

$$\frac{\Delta S_m}{R} = \left(1 - \frac{r-1}{r}\theta\right)^{-1} \left\{ -\frac{\theta}{r} \log \theta - (1-\theta) \log (1-\theta) \right\}. \quad (11.6)$$

By introducing the mole fraction x , this can be written in the simpler form

$$\frac{\Delta S_m}{R} = -x \log \theta - (1-x) \log (1-\theta). \quad (11.7)$$

We shall later refer to this as Flory's formula.

So much for open chain molecules. We know of only three previous attempts to extend these formulae to molecules containing a single closed ring. The simplest suggestion made by Miller (1947) is that formulae (11.4) and (11.5) should be used as they stand, merely giving q the value defined by (11.2) instead of (11.1). A different suggestion due to Staverman (1950) is that (11.4) should be replaced by a formula in which the quantity in brackets in the numerator is raised to the power r instead of $r-1$. Neither of these suggestions is confirmed by the formulae obtained in the present paper.

The other attack on these problems is of a different character. Mathot (1949) in his thesis considers the problem of placing square molecules on a planar square lattice. His method of approach is essentially the same as ours, and it was, in fact, his thesis which stimulated the present work. Mathot was, however, unfortunate in two respects. First, he did not properly understand the equivalence relations. He assumed wrongly a relation $\xi = \zeta^2$ between the quantities in table 4, and this false equivalence relation leads to contradictions. Secondly, he was unfortunate in choosing the example of squares on a planar lattice, for we have seen that the treatment of this particular example leads to peculiar inconsistencies.

12. NUMERICAL VALUES

We have used the formulae of earlier sections to calculate numerical values for the molar entropy of mixing ΔS_m . The results for trimers are given in table 5. The first column gives the volume fraction θ of trimers, the second column the corresponding value of the mole fraction x of trimers. The third column gives the values of $\Delta S_m/R$ for open chain trimers on a simple cubic lattice calculated by means of

TABLE 5. VALUES OF $\Delta S_m/R$ FOR MIXTURES OF TRIMERS WITH MONOMERS

θ	x	open chains	triangles	Flory's approximation	ideal solution
		$z = 6$	$z = 12$	$z = \infty$	
0.05	0.01724	0.0981	0.0981	0.1021	0.0871
0.1	0.03571	0.1760	0.1760	0.1838	0.1541
0.2	0.07692	0.3147	0.3147	0.3298	0.2712
0.3	0.1250	0.4410	0.4411	0.4626	0.3768
0.4	0.1818	0.5575	0.5576	0.5846	0.4741
0.5	0.2500	0.6618	0.6621	0.6931	0.5623
0.6	0.3333	0.7475	0.7479	0.7811	0.6365
0.7	0.4375	0.7998	0.8004	0.8333	0.6853
0.8	0.5714	0.7879	0.7885	0.8173	0.6829
0.9	0.7500	0.6352	0.6357	0.6547	0.5623
0.95	0.8636	0.4416	0.4419	0.4528	0.3983
0.98	0.9423	0.2398	0.2400	0.2447	0.2206
0.99	0.9706	0.1427	0.1427	0.1452	0.1327

TABLE 6. VALUES FOR $\Delta S_m/R$ FOR MIXTURES OF TETRAMERS WITH MONOMERS

θ	x	open chains	tetrahedra	squares	Flory's approximation	ideal solution
		$z = 6$	$z = 12$	$z = 6$	$z = \infty$	
0.05	0.01299	0.0844	0.0823	0.0812	0.0895	0.0693
0.1	0.02703	0.1547	0.1505	0.1483	0.1647	0.1243
0.2	0.05882	0.2850	0.2770	0.2727	0.3047	0.2237
0.3	0.09677	0.4071	0.3958	0.3895	0.4357	0.3150
0.4	0.1429	0.5290	0.5152	0.5069	0.5655	0.4068
0.5	0.2000	0.6502	0.6348	0.6250	0.6931	0.5004
0.6	0.2727	0.7583	0.7429	0.7318	0.8057	0.5860
0.7	0.3684	0.8434	0.8294	0.8180	0.8918	0.6581
0.8	0.5000	0.8720	0.8617	0.8512	0.9163	0.6931
0.9	0.6923	0.7505	0.7457	0.7384	0.7814	0.6173
0.95	0.8261	0.5448	0.5430	0.5386	0.5634	0.4620
0.98	0.9245	0.3056	0.3051	0.3032	0.3139	0.2676
0.99	0.9612	0.1842	0.1840	0.1830	0.1885	0.1642

formula (11.5) with $z = 6$, $r = 3$, $q = 7/3$. The fourth column gives the values of $\Delta S_m/R$ for triangular trimers on a spatial close-packed lattice calculated from formula (10.5). The fifth column gives the values of $\Delta S_m/R$ calculated from Flory's formula (11.7), and the sixth column the values of $\Delta S_m/R$ for an ideal solution at the same value of the mole fraction x . It will be noticed that in spite of the quite different forms of equations (10.5) and (11.5) there is practically no difference in the numerical values of $\Delta S_m/R$ between triangles and open chain trimers. It will also be noticed

that Flory's formula gives a much closer approximation to these values than does the formula of an ideal solution.

The results for tetramers are given in table 6. The first column gives the volume fraction θ of tetramer and the second column its mole fraction x . The third column gives the values of $\Delta S_m/R$ for open chain tetramers on a simple cubic lattice calculated by means of formula (11.5) with $z = 6$, $r = 4$, $q = 3$. The fourth and fifth columns give the values of ΔS_m for tetrahedra on a close packed lattice and squares on a simple cubic lattice calculated by formulae (10.7) and (10.9) respectively. The sixth column gives the values obtained from Flory's formula (11.7) and the seventh column the values in an ideal solution at the same value of the mole fraction x . The differences in the calculated values of $\Delta S_m/R$ between the three shapes of tetramers are small, but not so small as between the two kinds of trimers. Again it will be noticed that Flory's formula gives a considerably better approximation to these values than does the formula for ideal solutions.

13. COMPARISON WITH FLORY'S FORMULA

Since it is clear from tables 5 and 6 that both for trimers and tetramers Flory's formula gives a much better approximation to our values for the molar entropies of mixing than does the formula for ideal solutions, it seems worth while pursuing the comparison further. The comparison is shown more strikingly in figure 1 where the difference in $\Delta S_m/R$ between Flory's and our formulae is plotted against the mole fraction x . It will be seen that the greatest discrepancy is 0.075 which is about 10 % on the total entropy of mixing.

It has already been mentioned that Flory's formula is obtained formally from the formula for open chain molecules by making z tend to infinity. In fact if in (11.5) we expand the logarithms and substitute for $r - q$ the value given by (11.1) we obtain

$$\frac{\Delta S_m}{R} = \left(1 - \frac{r-1}{r}\theta\right)^{-1} \left\{ -\frac{\theta}{r} \log \theta - (1-\theta) \log (1-\theta) - \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left(\frac{2}{z}\right)^n \left(\frac{r-1}{r}\right)^{n+1} \theta(1-\theta^n) \right\}, \quad (13.1)$$

where the leading terms, namely, those not containing z , are just those of Flory's formula (11.6).

We shall now verify that Flory's formula is likewise obtainable by making z tend to infinity in our formulae for triangles, for tetrahedra and for squares. We have only to expand in powers of z^{-1} . The formulae obtained retaining terms in z^{-1} and neglecting terms in z^{-2} and higher powers are as follows:

Triangles:

$$\frac{\Delta S_m}{R} = (3-2\theta)^{-1} \left\{ -\theta \log \theta - 3(1-\theta) \log (1-\theta) - \frac{3z'-1}{zz'} \theta(1-\theta) - \dots \right\}. \quad (13.2)$$

Tetrahedra:

$$\frac{\Delta S_m}{R} = (4-3\theta)^{-1} \left\{ -\theta \log \theta - 4(1-\theta) \log (1-\theta) - \frac{9z'-9}{\frac{2}{3}zz'} \theta(1-\theta) - \dots \right\}. \quad (13.3)$$

Squares:

$$\frac{\Delta S_m}{R} = (4-3\theta)^{-1} \left\{ -\theta \log \theta - 4(1-\theta) \log (1-\theta) - \frac{4z'-1}{zz'} \theta(1-\theta) - \dots \right\}. \quad (13.4)$$

We observe that in all these formulae the leading terms, namely, those not containing z , are just those of Flory's formula. Thus Flory's formula is obtained formally from these formulae by letting z tend to infinity.

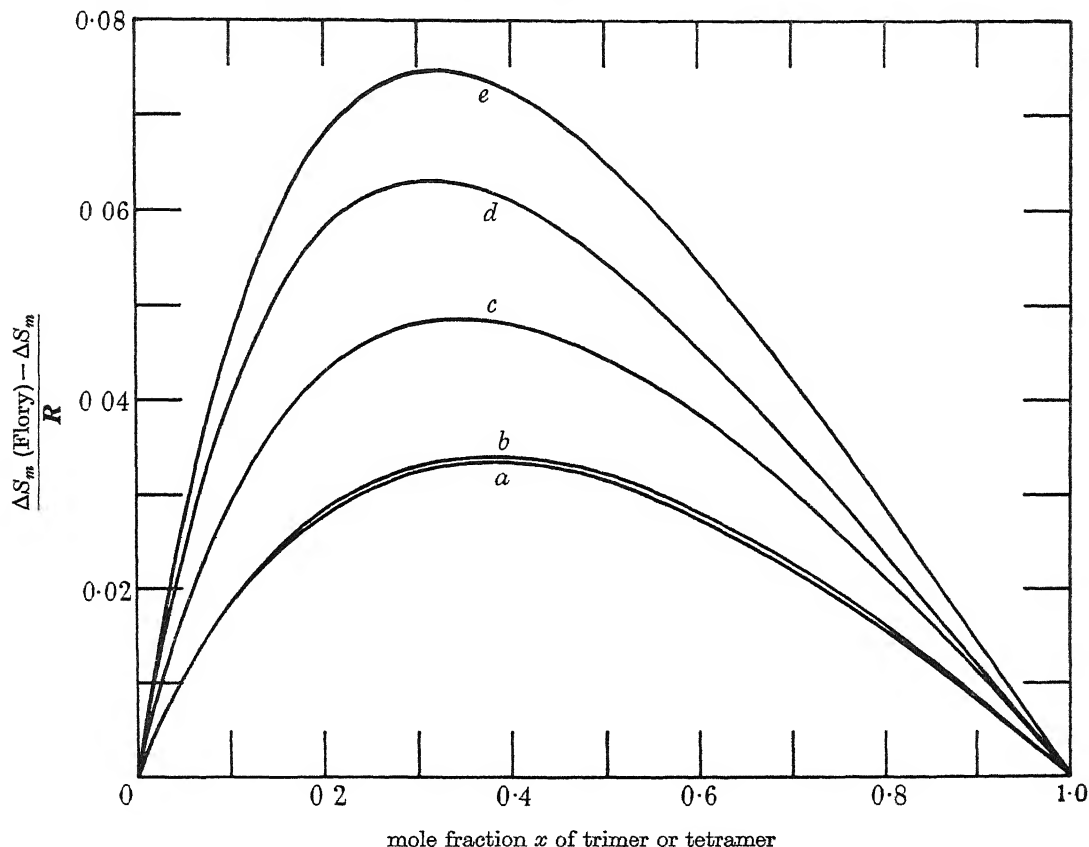


FIGURE 1. Difference in $\Delta S_m/R$ between Flory's and our formulae. Curve *a*, triangles; *b*, linear trimers; *c*, linear tetramers; *d*, tetrahedra; *e*, squares.

REFERENCES

- Chang, T. S. 1939*a* *Proc. Roy. Soc. A*, **169**, 512.
 Chang, T. S. 1939*b* *Proc. Camb. Phil. Soc.* **35**, 265.
 Flory, P. J. 1942 *J. Chem. Phys.* **10**, 51.
 Fowler, R. H. & Guggenheim, E. A. 1939 *Statistical thermodynamics*, p. 66. Cambridge University Press.
 Fowler, R. H. & Rushbrooke, G. S. 1937 *Trans. Faraday Soc.* **33**, 1272.
 Guggenheim, E. A. 1933 *Modern thermodynamics*, p. 103. London: Methuen.
 Guggenheim, E. A. 1937 *Trans. Faraday Soc.* **33**, 151.
 Guggenheim, E. A. 1944 *Proc. Roy. Soc. A*, **183**, 203.
 Guggenheim, E. A. 1945 *Trans. Faraday Soc.* **41**, 107.
 Guggenheim, E. A. 1949 *Thermodynamics*. North Holland Publishing Co.

- Huggins, M. L. 1942 *Ann. N.Y. Acad. Sci.* **43**, 9.
Mathot, V. 1949 *Propriétés thermodynamiques des solutions d'hydrocarbures*. Thesis, Brussels University.
Miller, A. R. 1942 *Proc. Camb. Phil. Soc.* **38**, 109
Miller, A. R. 1943 *Proc. Camb. Phil. Soc.* **39**, 54.
Miller, A. R. 1947 *Proc. Camb. Phil. Soc.* **42**, 422.
Miller, A. R. 1948 *Theory of solutions of high polymers*. Oxford: Clarendon Press
Staverman, A. J. 1950 *Rec. Trav. Chim. Pays-Bas*, **69**, 163 (seen in MS.).
-

A study of sensitized explosions.

VIII. Experimental work on the hydrogen-oxygen reaction sensitized by chloropicrin

BY P. G. ASHMORE AND R. G. W. NORRISH, F.R.S.

Laboratory of Physical Chemistry, University of Cambridge

(Received 9 May 1950)

The boundaries of ignition of hydrogen-oxygen mixtures in the presence of traces of chloropicrin have been investigated by the admission method. The experimental results show that the lower and upper sensitizer limits between 300 and 400° C, and their variations with pressure of reactants and with temperature, are determined by the nitrosyl chloride formed by the rapid decomposition of the chloropicrin. Estimates of the rate of flow of gases into the reaction vessel show that above 370° C the decomposition is completed within the entry tube. The induction periods are shorter than with nitrosyl chloride as sensitizer, and are increased without altering the position of the sensitizer limits either by heating a greater length of the entry tube or by gradually replacing chloropicrin by nitrosyl chloride. These effects are attributed to changes in the concentration of chain centres arriving in the vessel. The hypothesis is made that free chlorine atoms are formed during the decomposition of the chloropicrin and initiate the chains, unless removed by nitrosyl chloride. These chains decay in narrow tubes, but multiply under favourable conditions by branching reactions involving nitrosyl chloride to give branched chain-thermal ignitions.

It is suggested that additional support for this hypothesis would be obtained if chloropicrin were found to sensitize the hydrogen-chlorine reaction, because in this reaction the nitrosyl chloride formed by decomposition of the chloropicrin would act as an inhibitor.

INTRODUCTION

Previous papers in this series (Foord & Norrish 1935; Dainton & Norrish 1941) have shown that the sensitizing action of nitrogen peroxide and of nitrosyl chloride on the hydrogen-oxygen reaction can be ascribed to the provision of chain-initiating and chain-branching reactions by the sensitizer at temperatures (300 to 400° C) where the unsensitized reaction would proceed very slowly. At these temperatures the vapour of chloropicrin decomposes very rapidly to yield nitrosyl chloride and phosgene as main products. It seemed of interest to discover whether chloropicrin sensitizes the hydrogen-oxygen reaction, and if so, whether the sensitization is due to the nitrosyl chloride formed from the chloropicrin.

The present paper (part VIII) describes experimental work, begun in 1938 and completed after the war, on the sensitizing effect of chloropicrin vapour on the hydrogen-oxygen reaction. The behaviour of chloropicrin is attributed in part to the provision of chain-branching reactions by the nitrosyl chloride formed from the chloropicrin, and in part to the appearance during the decomposition of chloropicrin of a plentiful supply of free atoms which can readily initiate the hydrogen-oxygen chains. A consideration of previously published work on the decomposition of chloropicrin suggests that these atoms are chlorine atoms.

This hypothesis suggested that chloropicrin might sensitize the hydrogen-chlorine reaction. In addition to providing a test for the hypothesis, the lowering of the ignition temperature of a selected concentration of hydrogen and chlorine would be of interest, as the reaction is usually considered to proceed by straight chains. The hypothesis is fully substantiated by the work described in the following paper (part IX).

In a later paper (part X) the kinetics of decomposition of chloropicrin are discussed in relation to the above hypothesis and to previously published work on the decomposition of chloropicrin. Part X also gives an outline of the kinetics of the hydrogen-oxygen and the hydrogen-chlorine reactions sensitized by chloropicrin.

EXPERIMENTAL METHOD

Apparatus

The apparatus used resembled that of Dainton & Norrish (1941) and was evacuated by a two-stage quartz mercury diffusion pump backed by a 'Hyvac' rotary pump. The cylindrical reaction vessels were supported horizontally in a furnace whose temperature was controlled to $\pm 1^\circ \text{C}$ by a Cambridge Instrument Company regulator, operated by changes in the e.m.f. of a chromel-alumel thermocouple. The vessels used for most experiments were of Pyrex, 28 mm. in diameter. Vessel *A* had a plain entry tube of 6 mm. internal diameter. Vessel *B* had two entry tubes, each of the same length and of internal diameter 6 mm. The 'heated' entry tube ran the length of the furnace before entering the vessel, while the 'normal' entry tube was disposed as far as possible outside the furnace and could be heated by a small subsidiary furnace. Gas pressures were measured, according to their range, on a McLeod gauge, a glass spoon gauge sensitive to 0.02 mm. of mercury, or a mercury manometer. Apiezon grease M or L was used on stopcocks. This grease absorbs very small amounts of chloropicrin vapour, but the amount absorbed does not affect the reproducibility of results.

Preparation of gases

Hydrogen, oxygen and argon were prepared, purified and stored as described by Dainton & Norrish (1941).

Chloropicrin was obtained from a commercial stock (1939) and from C.D.E.S. (Porton) (1946). After purification, these samples yielded products which gave vapour pressure-temperature relationships in good accord with published values. Each sample was placed in an all-glass apparatus and distilled at reduced pressure

over phosphorus pentoxide. The distillate was fractionated twice *in vacuo*, the middle third of each fractionation being collected in a trap cooled in solid carbon dioxide. The chloropierin was stored in a trap in solid carbon dioxide.

Nitrosyl chloride was taken from a sample supplied by I.C.I. Ltd. and was fractionated *in vacuo* after passage over phosphorus pentoxide. In order to remove traces of chlorine, the nitrosyl chloride was distilled from -80°C (solid carbon dioxide-ether) and collected in a trap cooled to -110°C (a freezing mixture of CHCl_3 and $\text{CHCl}=\text{CCl}_2$ cooled with liquid nitrogen). The nitrosyl chloride was stored as the solid in a trap cooled in liquid nitrogen.

Carbonyl chloride was taken from a cylinder and fractionated *in vacuo*, after drying with phosphorus pentoxide. It condensed to a colourless liquid which was stored in a trap cooled in solid carbon dioxide.

Experimental technique

The reactants and sensitizer were premixed for 20 min. in the required proportions and admitted to the reaction vessel already heated to the required temperature. Small pressures of sensitizer were measured using the 'dilute mixture' technique described by Dainton & Norrish. All concentrations are given as mm. Hg at 400°C . The criterion of ignition was the visual observation of a hot flame, check experiments at various concentrations confirming that this was accompanied by nearly 100% combustion in the reaction vessel. At lower total pressures, the Bourdon gauge was left connected and a sharp kick (indicating an increase of pressure due to heating) marked the onset of ignition. Slow reactions were followed on the Bourdon gauge. The small 'kick' before a slow reaction was observed with nitrosyl chloride and with the decomposition products of chloropierin, but was never observed with chloropierin. Induction periods were measured to ± 0.2 sec. on a stop-watch, and where ignitions are reported as 'instantaneous' the induction periods were certainly less than 0.2 sec.

EXPERIMENTAL RESULTS

Mixtures of hydrogen and oxygen, to which have been added traces of chloropierin, ignite on admission to a reaction vessel at temperatures below 400°C , provided that the concentration of chloropierin lies between a lower and an upper limit. At 370°C , with 150 mm. of hydrogen and oxygen in the stoichiometric proportions, these limits have the values 0.12 and 0.83 mm. Just below the lower limit of ignition, the rate of reaction is fast (figure 1), and the reaction begins as soon as the gases are admitted. Just above the upper limit there is an appreciable induction period, which is followed by a comparatively slow reaction. The trends of the rates as the limits are approached are shown in figure 2. The induction periods near the lower limit are too short to be measured; with chloropierin concentrations above 0.5 mm. they increase in an approximately linear manner through and beyond the upper limit. Ignitions which occur between the limits are accompanied by a bright yellow flame and a loud click, and result in the complete conversion of the main reactants in the vessel into water.

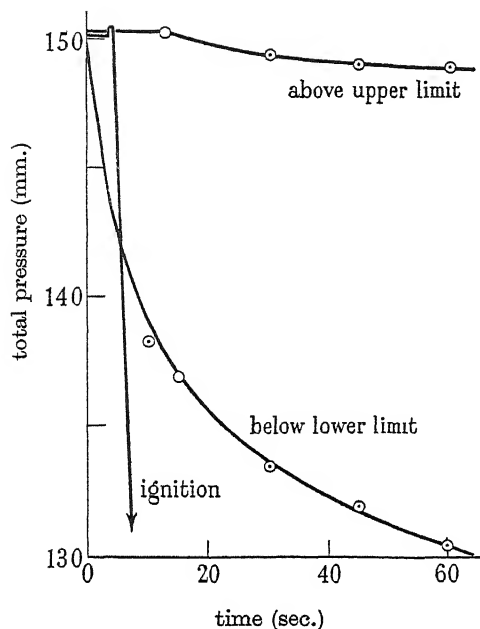


FIGURE 1. Typical pressure-time curves for 0.075 mm. of chloropicrin (below the lower limit of ignition), 0.75 mm. of chloropicrin (ignition), and 0.87 mm. of chloropicrin (above the upper limit of ignition). 150 mm. of $2\text{H}_2 + \text{O}_2$ at 370°C in Pyrex vessel A, 28 mm. internal diameter.

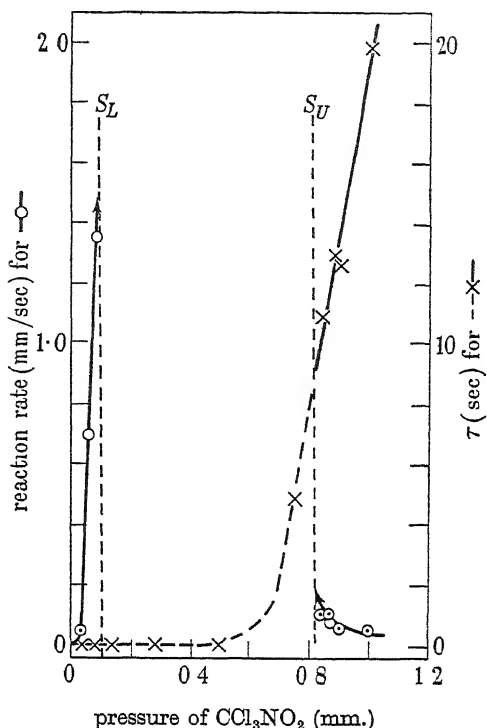


FIGURE 2. The variation of reaction rate (\odot) and induction period τ (\times) with pressure of chloropicrin added to 150 mm. of $2\text{H}_2 + \text{O}_2$ at 370°C in Pyrex vessel A. The lower (S_L) and upper (S_U) sensitizer limits of ignition. Broken lines indicate ignition.

Variation of sensitizer limits with temperature and total pressure

The variations of the lower and upper sensitizer limits with temperature were investigated for concentrations equivalent to 75, 150, 210, 300 and 450 mm. at 400°C . The results are shown in figure 3. The lower limit is relatively independent of temperature for all the concentrations except for 450 mm. at lower temperatures, whereas the upper limit is markedly dependent upon the temperature at all concentrations. The upper and lower limits converge as the temperature is lowered, and below a certain temperature there is no region of ignition. Instead, the reaction rate passes through a maximum as the concentration of chloropicrin is increased.

The variation of the lower and upper sensitizer limits with total pressure, at a constant temperature, is illustrated by figure 4 (for 374°C). The lower limit is almost independent of pressure and of the way the gases flow into the vessel. The upper limit changes considerably with the total pressure, the curve showing a break at about 200 mm. Below 200 mm., the values of the upper limits are independent of the way the gases flow into the vessel. Above 200 mm., ignitions near the upper limit are instantaneous, and the value of the upper limit is higher with a wider tap

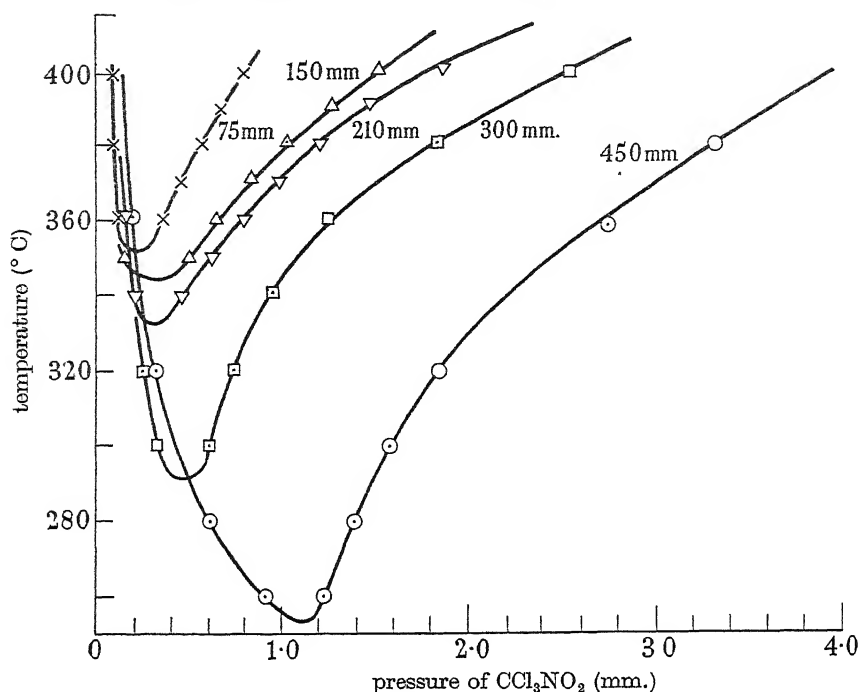


FIGURE 3. The dependence of the sensitizer limits of ignition (chloropicrin) upon temperature with different total pressures of $2\text{H}_2 + \text{O}_2$, in Pyrex vessel A. Ignition occurs within each curve.

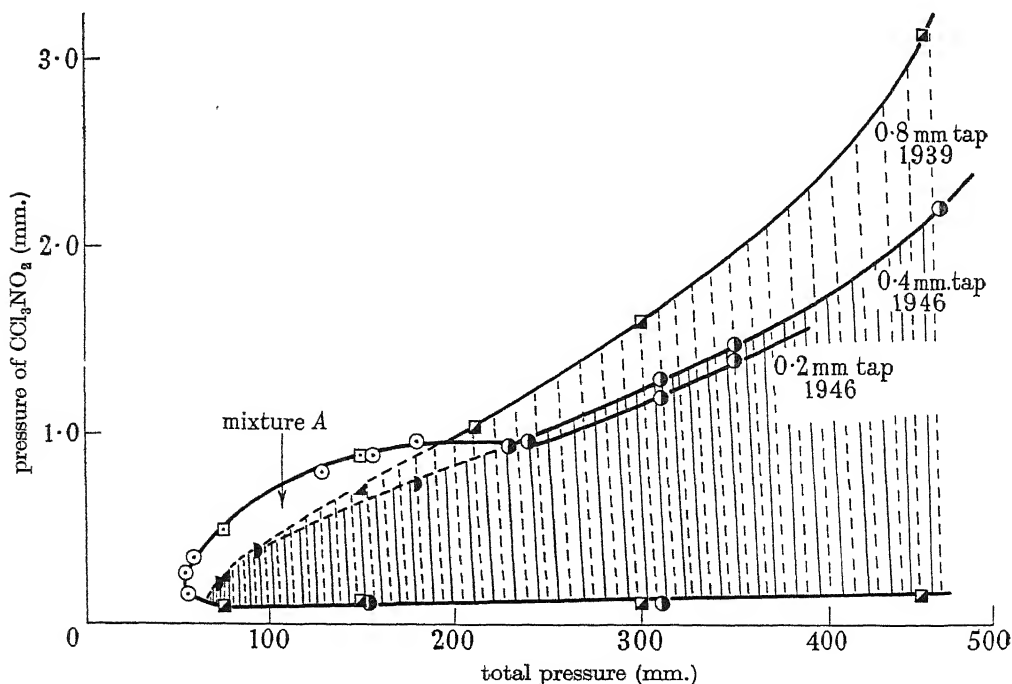


FIGURE 4. The dependence of the sensitizer limits (chloropicrin) upon the total pressure of $2\text{H}_2 + \text{O}_2$ at 374°C , in Pyrex vessel A. Ignition boundaries with 8 mm. tap (1939) shown (\square) and with 2 and 4 mm. taps (1946) shown (\circ). Regions of instantaneous ignition are shaded, and their boundaries are shown by the half-squares \blacktriangle and half-circles \bullet .

between mixing vessel and reaction vessel. If successively higher pressures of argon are added to a selected mixture of reactants and sensitizer which normally ignites (for example, *A* in figure 4), the induction periods at first fall slightly and the ignitions are eventually quenched. With further addition of argon, however, instantaneous ignitions can be observed.

Other experiments confirm that when the ignitions at the upper limit are instantaneous, the position of this limit depends upon the construction of the inlet tubes and taps used to admit the gases.

The induction periods

At a constant concentration of reactants, the induction periods are always short near the lower limit. If the concentration of reactants is below 200 mm., the induction periods are measurable near the upper limit, but are shorter at higher

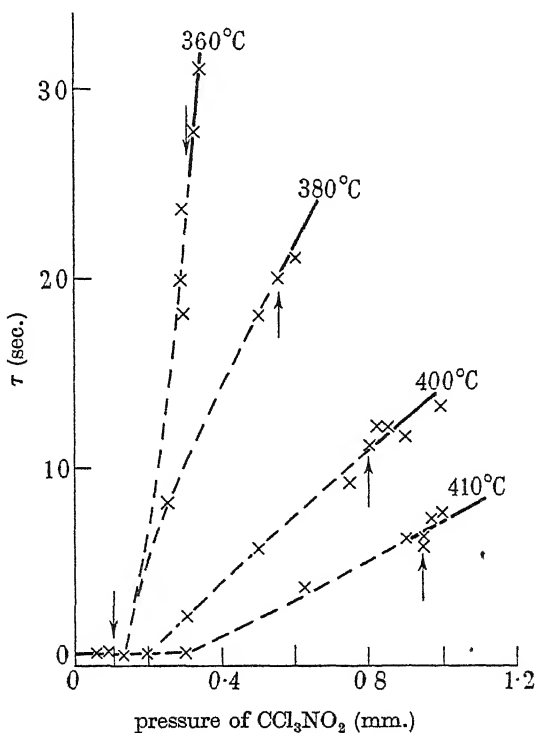


FIGURE 5. The relation between induction period τ and pressure of chloropicrin at various temperatures with 75 mm. of $2\text{H}_2 + \text{O}_2$. Pyrex vessel *A*. Broken lines indicate ignition.

temperatures as shown in figure 5. At higher concentrations of reactants, the induction period may be too small to be measured even at the upper sensitizer limit. Individual points on the curves of figure 5 are reproducible to within a few seconds for any particular construction of the tubes and taps admitting the gases to the furnace, the values of the induction periods near the upper limit are substantially altered, however, by a change in the time taken by the gases to pass through the heated portion of the narrow entry tube. This was demonstrated clearly by using

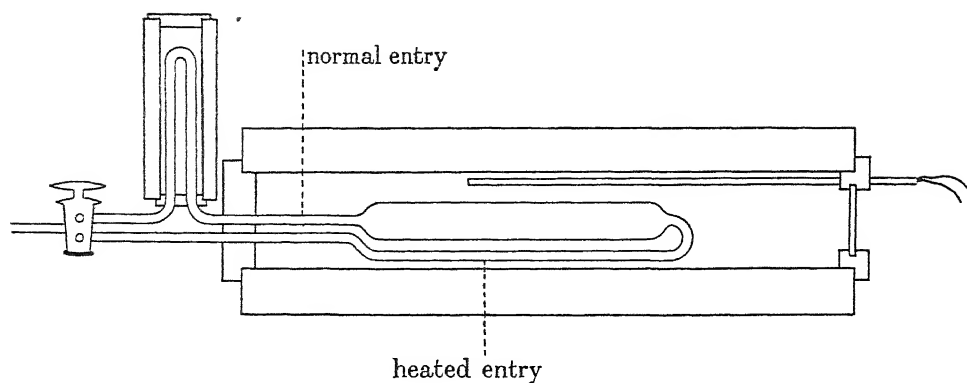


FIGURE 6. Pyrex vessel B, 28 mm. internal diameter, showing normal and heated entry.

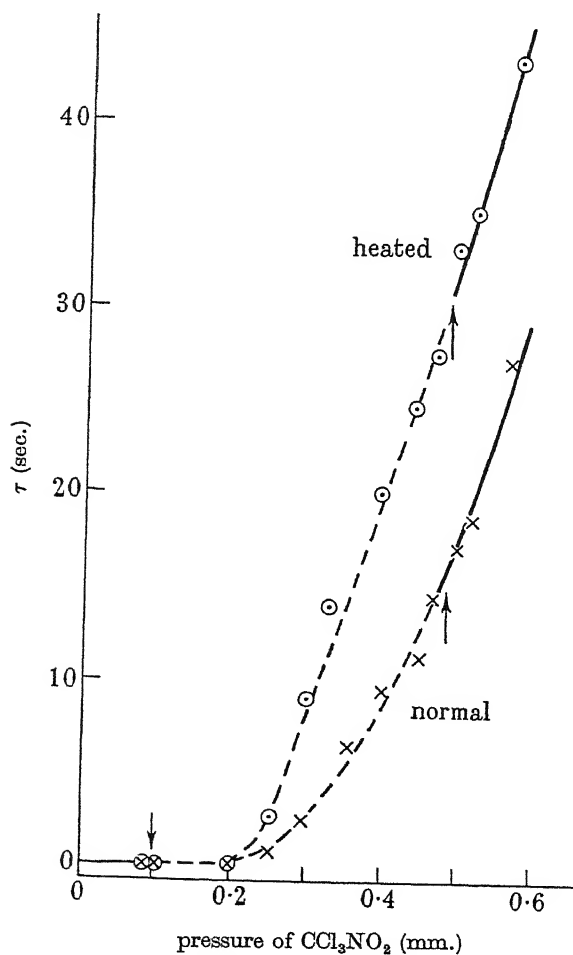


FIGURE 7. The dependence of the induction period τ upon pressure of chloropierin with two different lengths of heated entry tube. Results shown for the 'normal' (x) and 'heated' (o) tubes of Pyrex vessel B, with 90 mm. of $2\text{H}_2 + \text{O}_2$ at 374°C . Broken lines indicate ignition.

reaction vessel *B* (figure 6). The route was preselected by the two-way tap, and the gases were admitted by opening an 8 mm. tap in the mixing vessel. The velocity of flow of gases into the vessel should be the same for both routes. The induction

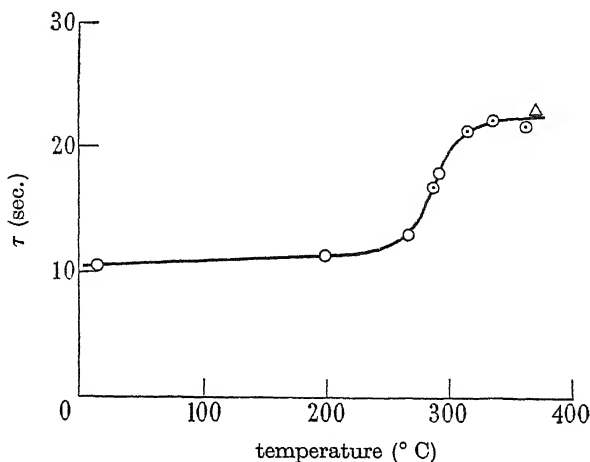


FIGURE 8. The change in induction period τ when the 'normal' entry tube of vessel *B* is heated with the subsidiary furnace (○). The corresponding induction period for the 'heated' tube of vessel *B* is shown (△). 90 mm. of $2\text{H}_2 + \text{O}_2$ at 374°C . 0.45 mm. CCl_3NO_2 .

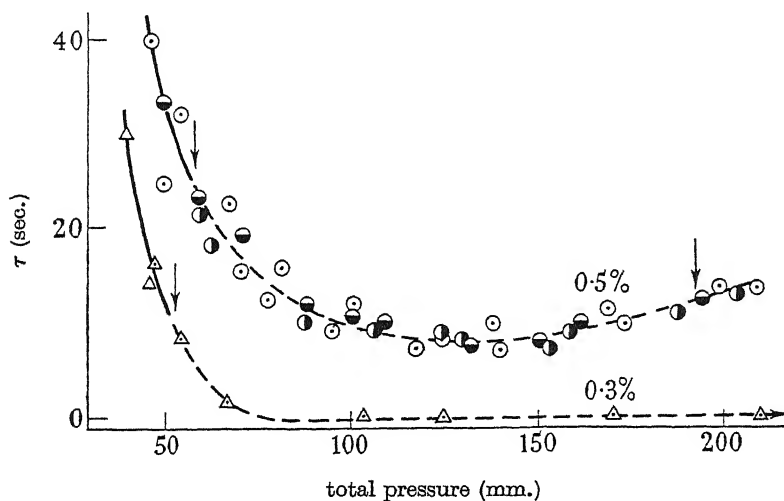


FIGURE 9. The dependence of the induction period τ upon the total pressure of $2\text{H}_2 + \text{O}_2$ at 374°C , in the presence of 0.3% and of 0.5% of chloropicrin. Pyrex vessel *A*. Broken lines indicate ignition. Age of dilute mixture: ○, 1 hr.; ●, 6 hr.; ◐, 20 hr.

periods were much longer when using the 'heated' entry tube, as illustrated for 90 mm. of reactants by the curves in figure 7. When the 'normal' tube was heated with the subsidiary furnace, the induction periods of a given mixture of reactants and sensitizer increased, and approached those observed for mixtures entering by the 'heated' tube (figure 8).

The change in the induction periods as the concentration of reactants is increased, keeping the percentage of chloropicrin constant, and the temperature constant, is illustrated by the curves in figure 9. This figure also shows that there is little change in the sensitizing action of chloropicrin which has been stored for some time in a dilute mixture with oxygen.

The sensitizing action of the decomposition products of chloropicrin

The decomposition products were prepared by passing chloropicrin vapour through a chamber heated to 370°C , and were stored as dilute mixtures with oxygen. The decomposition products, when used as sensitizer, gave lower and upper limits of ignition, outside which slow reactions occurred and between which ignitions were observed. The variation of these limits (expressed as actual pressure of de-

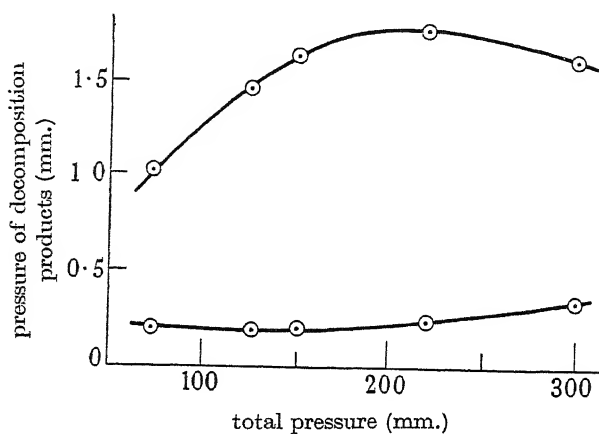


FIGURE 10. The dependence of the upper and lower sensitizer limits, using the decomposition products of chloropicrin, upon total pressure of $2\text{H}_2 + \text{O}_2$, at 374°C in Pyrex vessel A.

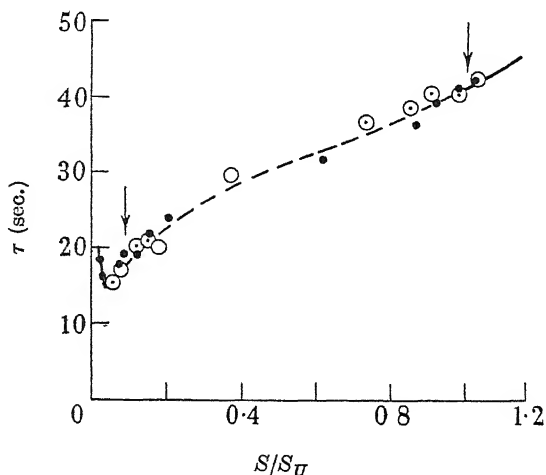


FIGURE 11. The dependence of the induction period τ upon the ratio S/S_U where S = pressure of sensitizer, S_U = upper sensitizer limit. Results shown (○) with nitrosyl chloride as sensitizer and (●) with the decomposition products of chloropicrin. 150 mm. of reactants at 374°C in Pyrex vessel A.

composition products) with the total concentration of reactants is shown in figure 10, and there is a strong suggestion of a closed curve for the ignition region. The values of the induction periods as the ignition region is crossed, at a constant total pressure, are very similar to those observed with nitrosyl chloride as sensitizer and are much longer than with chloropierin. In figure 11 the induction periods for

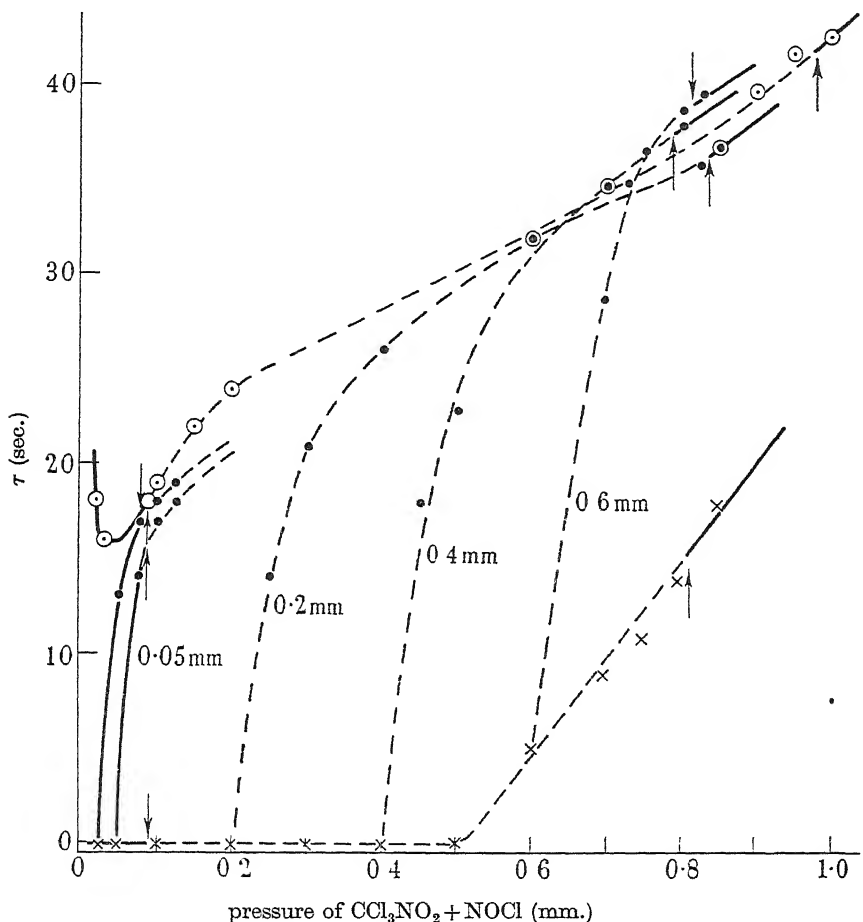


FIGURE 12. Changes produced in the induction period τ by adding nitrosyl chloride to various fixed amounts of chloropierin. Induction periods with nitrosyl chloride alone (\odot), with chloropierin alone (\times) and with the 'mixed' sensitizer (\bullet), for 150 mm. of $2\text{H}_2 + \text{O}_2$ at 374°C in Pyrex vessel A. Lower and upper sensitizer limits (nitrosyl chloride + chloropierin) shown by arrows on each curve.

the decomposition products and for nitrosyl chloride are plotted against S/S_U , where S is the concentration of either sensitizer and S_U the corresponding upper sensitizer limit.

A series of experiments with carbonyl chloride, at the concentrations in which it would be formed in the experiments with chloropierin, showed that by itself it has no sensitizing action on the hydrogen-oxygen reaction. Nor does the addition of small amounts of carbonyl chloride to mixtures sensitized by nitrosyl chloride have

any effect on the upper limit for ignition of nitrosyl chloride; the induction period is also unchanged.

Experiments were designed to use a sensitizer which simulated partially decomposed chloropicrin, on the assumption that nitrosyl chloride is the only product of the decomposition which can affect the reactions leading to ignition. It was found that if traces of nitrosyl chloride were added to the chloropicrin in making up the sensitized mixtures, the induction period became measurable over the whole range of ignitions and even below the lower limit of ignition. This is shown in figure 12 by

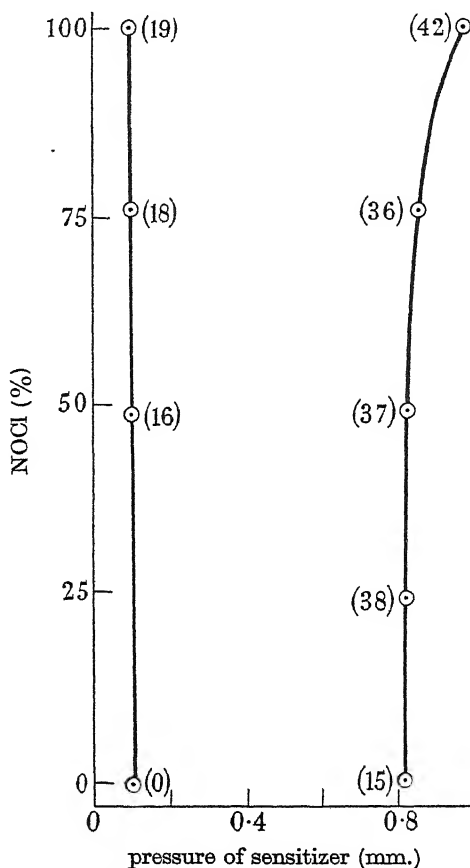


FIGURE 13. The dependence of the lower and upper sensitizer limits (chloropicrin + nitrosyl chloride) upon the percentage of nitrosyl chloride in the 'mixed' sensitizer. 150 mm. of $2\text{H}_2 + \text{O}_2$ at 374°C in Pyrex vessel A. Induction periods (sec.) in brackets.

the steeply rising curves, each curve showing the effect on the induction period of adding increasing amounts of nitrosyl chloride to a fixed amount of chloropicrin in making up the mixtures. Each of these steeply rising curves therefore springs from the point on the induction period curve for chloropicrin (crossed points) corresponding to the appropriate fixed amount of chloropicrin. Moreover, each steeply rising curve soon becomes practically coincident with the curve for nitrosyl chloride itself (circled points). The ignition limits are again indicated by arrows, and it can be seen that the lower limit occurs at the same total pressure of sensitizers on each

curve, while the upper limit increases somewhat for the curves rich in nitrosyl chloride. This effect can be seen more clearly in figure 13, in which the ignition limits are plotted as the total pressure of nitrosyl chloride plus chloropicrin against the percentage of nitrosyl chloride in the mixture. The induction periods (seconds) at the limits are shown by the numbers on the curves in figure 13.

True and 'shadow' boundaries of ignition

The appearance of an induction period of several seconds when traces of nitrosyl chloride are added to the chloropicrin has been utilized in experiments which indicate the nature of the upper chloropicrin limit at high pressures of reactants, for example, the upper limit above 200 mm. shown in figure 4. As soon as a trace of nitrosyl chloride is added to the chloropicrin, this limit no longer increases sharply with the increase of reactant pressure, but shows the same trend as with the sensitizer composed of the decomposition products of chloropicrin, figure 10, and as with nitrosyl chloride (Dainton & Norrish 1941). This remarkable change is illustrated in figure 14, which is the result of comparing (in the vessel *B*, using one entry tube only) the upper limit with 100% chloropicrin (circled points) and the upper limit with a sensitizer composed of 95% chloropicrin and 5% nitrosyl chloride (crossed points). The boundary of measurable induction periods with 100% chloropicrin is shown with solid semicircles. This boundary is lower if the flow of gases into the vessel is hindered by inserting a short length of capillary between the mixing vessel and the reaction vessel.

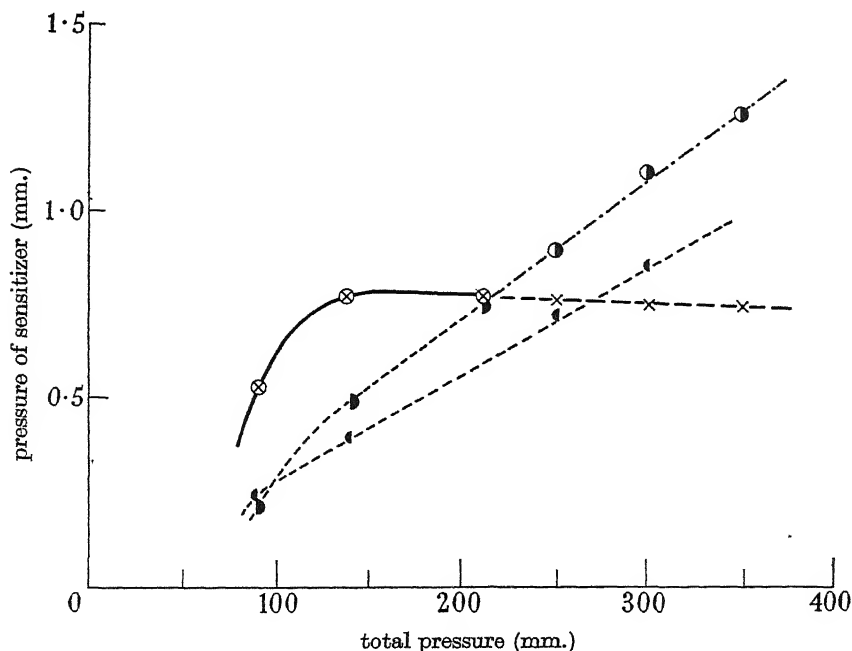


FIGURE 14. The dependence of the upper sensitizer limit upon the total pressure, with chloropicrin (O) and 95% chloropicrin + 5% nitrosyl chloride (x) at 374° C. The boundary of instantaneous ignition with chloropicrin is shown by half-circles (●) and the effect upon this boundary of inserting a capillary in the entry tube is shown by small half-circles (◐).

DISCUSSION OF RESULTS

The experimental work has established the following facts of major importance:

(a) Chloropicrin considerably extends the region of ignition of hydrogen-oxygen mixtures. At temperatures well below the minimum ignition temperature of a given pressure of the unsensitized reactants, ignitions occur with pressures of chloropicrin which lie between a lower and an upper sensitizer limit. Immediately outside these limits, slow reaction occurs.

(b) With a given total concentration of reactants, the lower sensitizer limit is almost independent of the temperature, but the upper sensitizer limit increases considerably as the temperature is raised.

(c) As the pressure of reactants is increased at a fixed temperature (say 370° C) the lower sensitizer limit changes little; the upper limit increases rapidly at low pressures (< 150 mm.), much less rapidly at moderate pressures (150 to 200 mm.), and then very rapidly at higher pressures.

(d) The upper limit at these higher reactant pressures is associated with an extremely short induction period, and the value found for the limit then depends upon the way in which the gases enter the vessel, being higher, for instance, with a wider inlet tap.

(e) The upper limit at lower reactant pressures (< 200 mm. at 370° C) is associated with an induction period of several seconds; the value of the limit is then independent of the way in which the gases enter the vessel, but the induction period is longer when a greater length of the entry tube is heated to furnace temperature.

(f) The induction periods near the lower limits are very short (a fraction of a second) at all temperatures and pressures, so short that the ignitions appear to be instantaneous

(g) When a trace of nitrosyl chloride is added to the chloropicrin, these very short induction periods are increased to several seconds. The lower limit is not changed. The upper limit is unaffected at reactant pressures below 200 mm., but above 200 mm. it no longer increases with increasing reactant pressure, but falls slightly.

(h) The gradual replacement of chloropicrin by nitrosyl chloride has no effect on the lower sensitizer limit (expressed as $P_L = P_{\text{CCl}_3\text{NO}_2} + P_{\text{NOCl}}$) and very little effect on the upper sensitizer limit until the proportion of nitrosyl chloride approaches 100 %. The induction periods increase rapidly in the early stages of the replacement, rapidly approaching values associated with 100 % nitrosyl chloride.

(i) When the decomposition products of chloropicrin are used as sensitizer, the limits are roughly double the corresponding limits with nitrosyl chloride or chloropicrin as sensitizer. The induction periods at the limits for the decomposition products are very similar to those at the corresponding limits for nitrosyl chloride

(j) Carbonyl chloride has no sensitizing action at the concentrations and temperatures at which chloropicrin and nitrosyl chloride are effective. Addition of carbonyl chloride has no effect on the upper limits of the nitrosyl chloride sensitized reaction, nor on the induction periods observed there.

The most important conclusion which can be drawn from these facts is that the behaviour of chloropicrin as a sensitizer closely resembles that of nitrosyl chloride. The resemblance is most marked in the close agreement between the values of the lower and the upper limits for the two sensitizers at moderate pressures of reactants. The fact that chloropicrin can be replaced by nitrosyl chloride with so little effect on the limits (figure 13), coupled with the evidence that carbonyl chloride (which is formed with nitrosyl chloride by the decomposition of chloropicrin) has no sensitizing action, strongly suggests that the sensitizer limits in the chloropicrin sensitized reaction are fixed by the nitrosyl chloride formed by the breakdown of the chloropicrin.

This hypothesis is supported by the agreement between the values of the sensitizer limits with chloropicrin and with the decomposition products, when the latter are expressed in terms of the nitrosyl chloride present (figure 15).

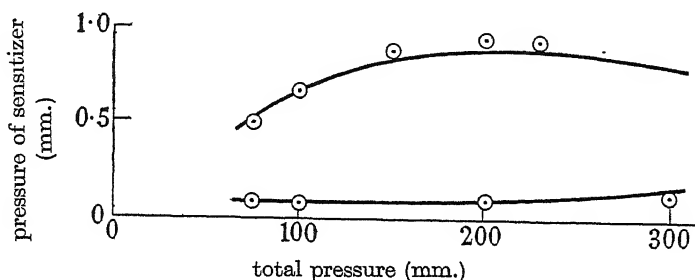


FIGURE 15. Comparison of the sensitizer limits using chloropicrin (\odot) and the decomposition products of chloropicrin (full lines, expressed as pressure of nitrosyl chloride present in the decomposition products) at various total pressures. 374°C in Pyrex vessel A.

It is also supported by the similar variation with temperature of the upper sensitizer limits with chloropicrin and with nitrosyl chloride. Dainton & Norrish (1941) showed that the upper limit P_U for nitrosyl chloride, with a constant concentration of reactants, fit the equation $\ln P_U = -E/RT + B$, where E varies somewhat with the total pressure of reactants as shown below. Figure 16 shows that the upper limits S_U for chloropicrin obey a similar equation, for $\log_{10} S_U$ is a linear function of the reciprocal of the absolute temperature. The values of E obtained from these lines compare well with those obtained by Dainton & Norrish for NOCl.

	concentration of reactants (mm.)		
	75	150	210
E chloropicrin (cal.)	17,100	18,300	18,600
E nitrosyl chloride (cal.)	13,600*	18,300	21,000

* The points which were used to determine this value were somewhat scattered.

Although the upper limit with chloropicrin apparently increases at higher pressures of reactants (figure 4), whereas it falls with the decomposition products (figure 10) and with nitrosyl chloride, the experiments illustrated in figure 14 suggest that when the very short induction period near the lower limit is lengthened the discrepancies between the upper limits with the different sensitizers disappear.

With 100 % chloropicrin, the upper ignition boundary at high reactant pressures is probably a projection of points on the boundary of 'instantaneous' ignitions; all mixtures whose composition passes through the region of extremely short induction periods during the entry of gases will ignite. The boundary observed will then be a 'shadow' of the region of very short induction periods. The boundary with 95 % chloropicrin + 5 % nitrosyl chloride, on the other hand, is probably the true upper boundary. The true trend of the upper limits with total pressure is therefore similar for chloropicrin and for nitrosyl chloride.

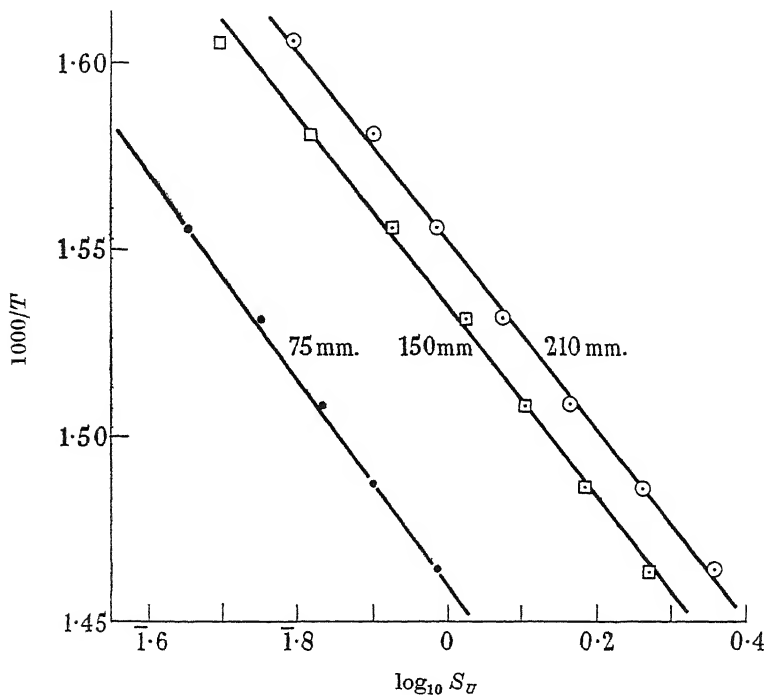


FIGURE 16. Variation of upper sensitizer (chloropicrin) limit with temperature: plots of $\log_{10} S_v$ against $1000/T$ for various total pressures of $2H_2 + O_2$, in Pyrex vessel A.

The discussion has shown that the sensitizing action of chloropicrin is so similar to that of nitrosyl chloride, especially in the experimental agreement between the values of the limits for the two sensitizers and the trends of these limits with reactant concentration and temperature, that it appears that nitrosyl chloride is the effective sensitizer in each system.

This hypothesis is supported by estimates of the percentage decomposition of the chloropicrin which occurs as the gases flow into the vessel. The minimum time spent by the gases in the entry tube can be calculated by treating the admission tap as a jet or nozzle, the bore of the tap forming the throat of the jet. Assuming that the flow is streamlined, the mass flow of gas through the tap can be calculated from the conditions of pressure and temperature in the mixing vessel, and hence, from the ratio of the volumes of the mixing vessel and reaction vessel, the 'ideal' time taken for the mixing vessel pressure to fall to the equi-

librium pressure. This 'ideal' time gives a minimum value for the actual time of flow, because the flow will certainly be turbulent and the calculations underestimate the effect during the later stages of flow of the back pressure built up in the reaction vessel. With the mixing and reaction vessels used, the minimum time for the pressure to equilibrate is about 100 msec. for a 2 mm. diameter tap and about 4 msec. for a 100 mm. diameter tap. It is also possible to calculate the ideal linear velocity of flow down the entry tube; with an entry tube of 6 mm. internal diameter, the average times spent by each small portion of the gases in the heated

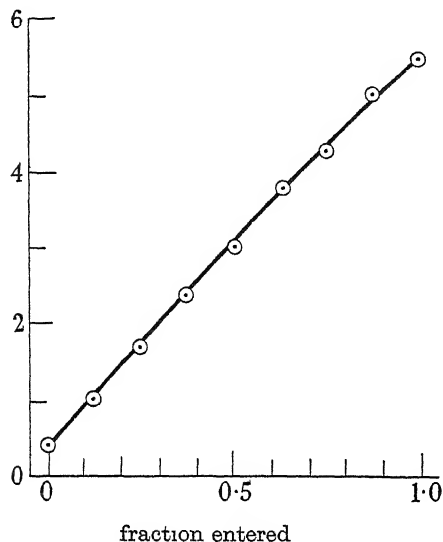


FIGURE 17. The calculated minimum time (msec.) spent by a portion of the gas in an entry tube 20 cm. long and 6 mm. internal diameter, at different stages of the flow into the vessel.

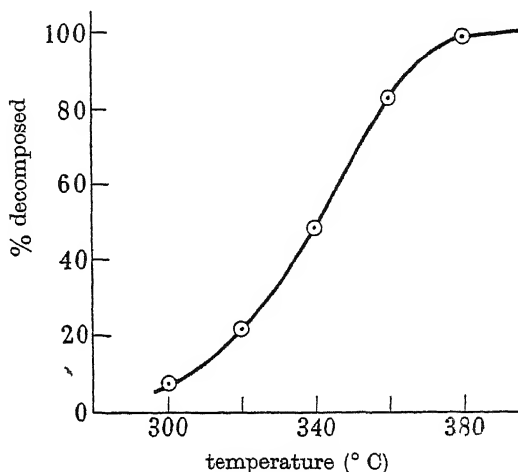


FIGURE 18. The estimated percentage of chloropicrin which has decomposed in 3 msec. at various temperatures.

portion (20 cm. long) can therefore be estimated (figure 17) at different stages of admission. The average time spent in the tube can be taken as roughly 3 msec. For various temperatures, the values of the velocity constant k of the unimolecular decomposition of chloropicrin have been calculated from the formula

$$k = 4.9 \times 10^{15} \times e^{-37,670/RT} \quad (\text{Smith \& Steacie 1938a, b}),$$

and hence the percentage decomposition in 3 msec. obtained (figure 18). At temperatures above 375°, the decomposition is virtually completed within the entry tube. It will certainly be completed within the much longer time taken to establish pressure equilibrium in the vessels.

There remains the problem why the induction periods are so much shorter with chloropicrin than with the same quantities of nitrosyl chloride, in spite of the fact that the position of the limits is fixed for both sensitizers by nitrosyl chloride. The shortening must be intimately connected with the rapid decomposition of the chloropicrin. Moreover, it is clear from figure 7 that the induction periods are longer when a greater length of the entry tube is heated. This suggests that short

induction periods result when the decomposition of the chloropicrin occurs close to the reaction vessel, under these circumstances, there would be a greater concentration of chloropicrin close to the vessel. The remarkable lengthening of the induction periods when a trace of nitrosyl chloride is added, however, eliminates chloropicrin as the *direct* cause of the shortening, because no interaction between chloropicrin and nitrosyl chloride has been observed during the decomposition of chloropicrin. The short induction periods must therefore be due to some reactive product of the decomposition of chloropicrin, probably an atom or free radical, which is removed rapidly by reaction with nitrosyl chloride; during the flow of chloropicrin-sensitized mixtures down the entry tube, the reactive product would be removed by the nitrosyl chloride formed from the chloropicrin, and its concentration would fall during the later stages of the flow. Assuming that the reactive product reacts with hydrogen or oxygen to initiate the hydrogen-oxygen chains, it is probable that most of the chains started in the entry tube would be started close to the point where the decomposition of the chloropicrin begins. If this point is remote from the reaction vessel in the narrow entry tube, the chains will not multiply, but rather decrease as the gases flow towards the vessel (because of the effective termination of chains on the walls of narrow tubes); once in the vessel, however, conditions may be favourable for the multiplication of the chains by branching, and ignition would then occur after a certain induction period. On the other hand, if the point lies close to the reaction vessel, the decrease in the number of chains would be less, in the vessel, the same multiplication will operate on a much greater initial number of chains, and the induction periods will be shortened. An extreme example of this shortening is provided by experiments with two quartz reaction vessels. These had wide entry tubes (10 mm. internal diameter) which were in effect elongated reaction vessels, and were favourable for multiplication of chains within the entry tube. It was observed that ignitions were instantaneous for all sensitized mixtures

The experimental results therefore lead to the conclusion that the induction periods are shorter with chloropicrin than with nitrosyl chloride because of the formation during the decomposition of chloropicrin of a reactive substance which can initiate hydrogen-oxygen chains, and which reacts rapidly with nitrosyl chloride. The rapidity of the reactions of this reactive substance suggest that it may be a free radical or atom.

CONCLUSIONS

The conclusions which have been reached may be recapitulated as follows:

(1) A careful comparison of the sensitizing action of chloropicrin and of nitrosyl chloride on the hydrogen-oxygen reaction suggests that nitrosyl chloride is the effective sensitizer in each system. In the chloropicrin-sensitized system, it is the decomposition product, nitrosyl chloride, which determines the values of the sensitizer limits and their variation with temperature and pressure.

(2) The velocity of decomposition of chloropicrin at temperatures above 370°C is probably sufficiently rapid for more than 90% of the decomposition to occur during the flow down a normal entry tube, and certainly sufficiently rapid for

complete decomposition to occur before the final pressures are reached in the reaction vessel.

(3) The difference between the induction periods of the reaction sensitized by chloropicrin and by the same quantities of nitrosyl chloride is probably due to the formation of a free radical or atom during the decomposition of chloropicrin. This free radical or atom (X) can initiate the hydrogen-oxygen chains, and is removed by nitrosyl chloride. This removal will occur as the sensitized mixtures pass down the entry tube into the reaction vessel, so that the concentration of X (and therefore the rate of initiation of the hydrogen-oxygen chains) is probably at a maximum in the entry tube just inside the furnace.

These conclusions would be consolidated if some other test for the occurrence of the free radical or atom X could be devised. A consideration of the structure of chloropicrin suggested, in the light of conclusion (3), that the radical or atom might be CCl_3 , COCl or $\text{Cl}\cdot$ * Such radicals would start the hydrogen-oxygen chains by reacting with hydrogen, and so would be expected to start other chain reactions involving hydrogen, for example, the hydrogen-chlorine reaction. This reaction is particularly interesting, because the nitrosyl chloride formed by the decomposition of chloropicrin is an inhibitor for chain reactions involving chlorine atoms, and so cannot play the same role as it does in the hydrogen-oxygen reaction. Any sensitization of the hydrogen-chlorine reaction must therefore be ascribed to chloropicrin or to the radical or atom formed by the decomposition of chloropicrin. It has been found that chloropicrin acts as a very powerful sensitizer for the hydrogen-chlorine reaction, and the detailed investigations described in the following paper fully confirm the existence of the free radical or atom X postulated in conclusion (3).

The authors wish to thank the Governing Body of Emmanuel College, Cambridge, for a Research Studentship given in 1938 to one of us (P.G.A.), and for a Research Fellowship (P.G.A.) which enabled this work to be resumed in 1946. They also wish to thank the Department of Scientific and Industrial Research for a Research Maintenance Grant in 1938 (P.G.A.).

REFERENCES

- Danton, F. S. & Norrish, R. G. W. 1941 *Proc. Roy. Soc. A*, **177**, 393, 411, 421.
Foord, S. A. & Norrish, R. G. W. 1935 *Proc. Roy. Soc. A*, **152**, 196.
Smith, W. McF. & Steacie, E. W. R. 1938a *J. Chem. Phys.* **6**, 145.
Smith, W. McF. & Steacie, E. W. R. 1938b *Canad. J. Res.*, **168**, 1, 222.

* In part X it will be shown that the radicals are probably chlorine atoms.

A study of sensitized explosions.

IX. Experimental work on the sensitization of ignitions of hydrogen-chlorine mixtures by chloropicrin

By P. G. ASHMORE AND R. G. W. NORRISH, F.R.S.

Laboratory of Physical Chemistry, University of Cambridge

(Received 9 May 1950)

The results of a study of the sensitizing action of chloropicrin upon the hydrogen-oxygen reaction suggested that chloropicrin might also sensitize the hydrogen-chlorine reactions by providing chlorine atoms which would initiate the reaction chains.

Experimental work is described which shows that the addition of small proportions of chloropicrin lowers the normal ignition boundary of hydrogen-chlorine mixtures to the temperature range 100 to 200° C. This sensitized boundary separates exceedingly rapid and complete reactions from slow reactions which are further retarded during their course by the formation of nitrosyl chloride from the chloropicrin. At higher temperatures (330 to 370° C) the decomposition of the chloropicrin is sufficiently rapid for the inhibiting action of the nitrosyl chloride to become prominent during the entry of the gases, and the ignition boundary rises with increasing temperature. It rises more sharply with a greater proportion of chloropicrin in the mixtures entering the vessel, and when a longer portion of the entry tube is heated. The sensitized ignitions are suppressed by the addition of small quantities of nitrosyl chloride. A remarkable 'backfire' phenomenon, in which ignitions travel back to the cold mixing vessel but not into the hot reaction vessel, is described and shown to be in agreement with the assumption that reaction chains are started by chlorine atoms formed from chloropicrin and are abruptly terminated by nitrosyl chloride also formed from chloropicrin.

Experimental work upon the effects of nitrosyl chloride upon the boundary of ignition of mixtures of hydrogen and chlorine show that below 400° C it acts as an inhibitor, but above 400° C it has a slight sensitizing effect.

INTRODUCTION

There has been comparatively little work on the thermal reaction between hydrogen and chlorine. The slow reaction has been studied by Pease (1934) between 200 and 230° C, by Morris & Pease (1939) at 184° C, by Christiansen (1929) at 200° C, and by Khodshaian & Kornfeld (1937) at 200° C. There is general agreement that the thermal reaction is, like the photochemical, a chain reaction, but at temperatures below 300° C the chains both start and end on the walls.

At higher temperatures the reaction proceeds sufficiently rapidly for inflammation to occur. The region of ignition was investigated by Sagulin (1928) who found that between 280 and 400° C there was a minimum ignition pressure (P) for each temperature (T). This pressure varied with the proportion of chlorine in the mixture admitted to the heated vessel, and, for a mixture of fixed proportions, with the diameter of the vessel. Sagulin found that there was a relation between P and T of the form

$$\log_{10} \frac{P}{T} = \frac{A}{T} + B,$$

where B varies with the vessel diameter and with the proportions of reactants used, but A is independent of these factors. These results were examined by Semenov (1935), who interpreted them as favouring a chain reaction initiated by chlorine

atoms formed in the gas phase by dissociation of chlorine, but not excluding a simple bimolecular reaction of plausible (but undetermined) activation energy.

It appeared from a study of the sensitizing action of chloropicrin upon the thermal reaction between hydrogen and oxygen that the decomposition of chloropicrin produces a radical or atom which would be capable of starting the hydrogen-chlorine chains. Moreover, the main products of the decomposition of chloropicrin, carbonyl chloride and nitrosyl chloride, would not be expected to initiate chains at temperatures below 300° C. Nitrosyl chloride is known to terminate chains involving chlorine atoms by reacting with them to give nitric oxide and chlorine (Dainton 1947). Carbonyl chloride and hydrogen are known to react photochemically, but the initiation step involves the breaking of a C-Cl bond and in the absence of light this step would be very slow at the temperatures concerned. It therefore appeared that the long chains of the hydrogen-chlorine reaction might be used as a magnifying agent to detect the presence of the atom or radical formed during the decomposition of chloropicrin.

The work described in this paper shows that chloropicrin sensitizes hydrogen-chlorine reaction to a remarkable extent, lowering the region of ignition by some two hundred degrees Centigrade with moderate pressures of reactants (10 to 200 mm. Hg). The detailed study of this effect has provided some striking illustrations of the interplay between sensitization by chloropicrin (by means of the atom or radical formed during its decomposition) and inhibition by the nitrosyl chloride which is also a product of decomposition of the chloropicrin. The results strongly support the suggestions advanced to explain the action of chloropicrin upon the hydrogen-oxygen reaction, and provide an interesting example of the sensitization of the hydrogen-chlorine reaction by a fragment of a molecule other than the reactants.

EXPERIMENTAL METHODS

Apparatus

The apparatus used was similar in most respects to that used for investigating the hydrogen-oxygen reaction, part VIII of this series (Ashmore & Norrish 1950a), except that all vessels were shielded from light and all connecting tubing and taps were enamelled black. Precautions were also taken to prevent stray light entering the observation window on the furnace during the admission of the gases and the subsequent reaction.

Two cylindrical Pyrex reaction vessels were used, each 28 mm. internal diameter, vessel *A* having a plain entry tube, and vessel *B* being the double entry vessel described in part VIII.

Preparation of gases

Hydrogen, nitrosyl chloride, carbonyl chloride, and chloropicrin were prepared and stored as described in part VIII.

Chlorine was prepared (a) by the method used by Ritchie & Norrish (1933). Pure hydrochloric acid was dropped on to pure potassium permanganate. The gas evolved was washed with aqueous permanganate and then with water, in darkened vessels, passed over phosphorus pentoxide, and collected in a trap in liquid nitrogen.

Any air trapped with the solid chlorine was pumped off with frequent melting and freezing of the chlorine. The liquid chlorine was then repeatedly distilled in a vacuum apparatus from -97°C (methylene chloride, solid-liquid) into liquid air, the middle two thirds being collected each time, and the final product being stored in a trap in liquid air. (b) From a cylinder of liquid chlorine, by passing the gas over phosphorus pentoxide and then collecting the chlorine in liquid air, and fractionating as above. Three fractionations were used. The chlorine prepared by each of these methods appeared to give the same ignition limits when comparison determinations were made.

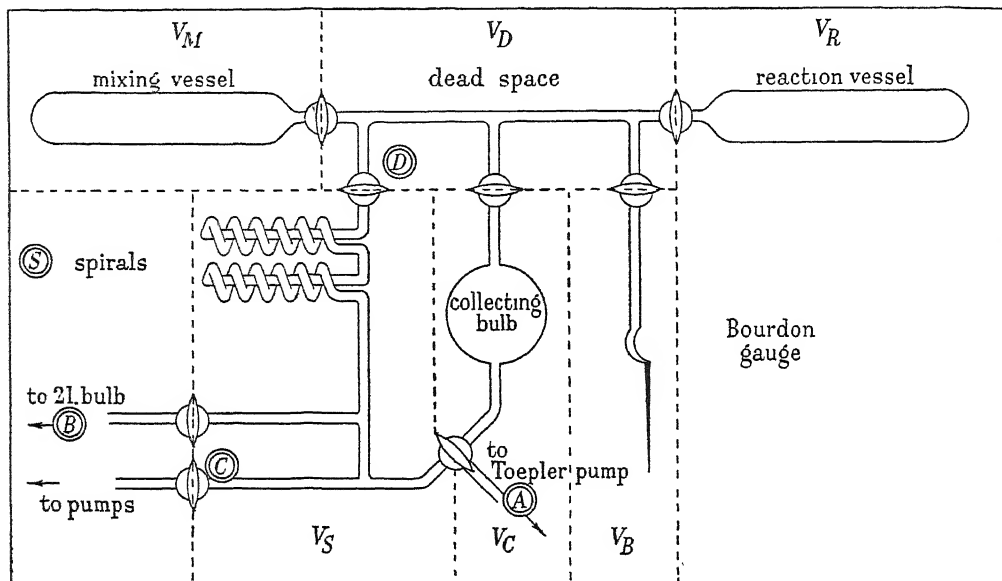


FIGURE 1. Apparatus for analysis of reacted mixtures.

Experimental technique

Ignition regions were explored using the admission method described in part VIII. The gas mixture was always admitted by opening the mixing vessel tap; when the double entry vessel was used, the desired route was first selected with the two-way tap. Ignitions were generally viewed through an observation window at the back of the furnace, but for some of the ignitions the peculiar phenomena encountered required simultaneous observation of the reaction vessel, connecting tubing and mixing vessel.*

Analysis of products

For the analyses during and after reaction, the apparatus shown in figure 1 was used. Gases in the reaction vessel could either be expanded through the spirals S , immersed in liquid nitrogen, into the 2 l. bulb B , or withdrawn through the spirals by the Toepler pump or by the main line pump by manipulation of the appropriate

* The authors are grateful to Dr A. J. C. Nicholson, Dr A. J. Harding, Mr E. E. Wilson, and Mr D. Patanyak for their assistance in these observations.

taps. The condensible gases (chlorine, hydrogen chloride, and chloropicrin and its decomposition products) were thus collected in the spirals and the hydrogen passed into the bulb or to the pump. The amount of hydrogen could only be determined by a lengthy withdrawal of the gases with the help of the Toepler pump. Trial runs, however, showed that a reliable figure for the percentage of reaction completed could be obtained from the amount of chlorine and hydrogen chloride retained in the spirals. This permitted a much more rapid determination to be made by opening the reaction vessel to the spirals and the two litre bulb, which reduced the pressure in the reaction vessel to less than 3% of its original value in a small fraction of a second, and the remaining gases could be rapidly withdrawn, again through the spirals, by opening the tap to the main pump. The absence of hydrogen chloride and of chlorine in the gases passing into the two litre bulb, was proved by withdrawing them slowly through the cooled spirals after the main bulk of the hydrogen chloride and the chlorine had been measured and pumped away. The amounts of gases withdrawn from the reaction vessel were measured as pressures in certain parts of the apparatus, and referred to the reaction vessel by using the measured ratios of the volumes of these parts to that of the reaction vessel.

EXPERIMENTAL RESULTS

Ignitions of unsensitized mixtures

The ignition region of equimolar mixtures of hydrogen and chlorine was investigated, using the Pyrex reaction vessel of 28 mm. internal diameter with the single entry tube. At a given temperature, ignition occurs with pressures above a limit, and the variation of this pressure with temperature gives a lower boundary to the region of ignition (figure 2). In order to obtain reproducible results, it was essential to use on the taps grease which had not been exposed to chloropicrin vapour. The ignitions in the region immediately above the boundary are sharp, that is, there is a hot bright flame accompanied by an audible click, and a sharp temporary pressure rise is registered on the Bourdon gauge when connected at lower pressures. The bright yellowish white flame strikes back into the mixing vessel at higher pressures. There is no observable time lag between admission and ignition.

The effect of chloropicrin

At 250° C, addition of 1 mm. of chloropicrin reduces the ignition pressure from above 250 mm. to below 20 mm. The new position of the ignition boundary in the presence of 2, 10 and 18 mm. of chloropicrin (expressed at 400° C) was determined, and the values obtained are also shown in figure 2.

Above 200° C the ignition boundary is very low even for small concentrations of chloropicrin. It was expected, however, that this boundary might rise again at higher temperatures, and accordingly the boundary for 0.55 mm. of chloropicrin was determined over the temperature range of 150 to 400° C, using the double entry vessel *B*. In figure 3 the sensitized boundary is compared with the boundary for unsensitized equimolar mixtures of hydrogen and chlorine. It will be seen that

the complete region of sensitized ignitions is shaped rather like the 'cold-flame' region found in the combustion of some ethers and hydrocarbons.

The position of the sensitized boundary between 340 and 360° C is greatly affected by the location of the entry tube inside the furnace and by the time taken for the gases to flow through it into the reaction vessel. Figure 4 shows the two boundaries obtained with the heated and normal entry tubes of vessel *B*, and also the effect of inserting a capillary constriction in the entry tube.

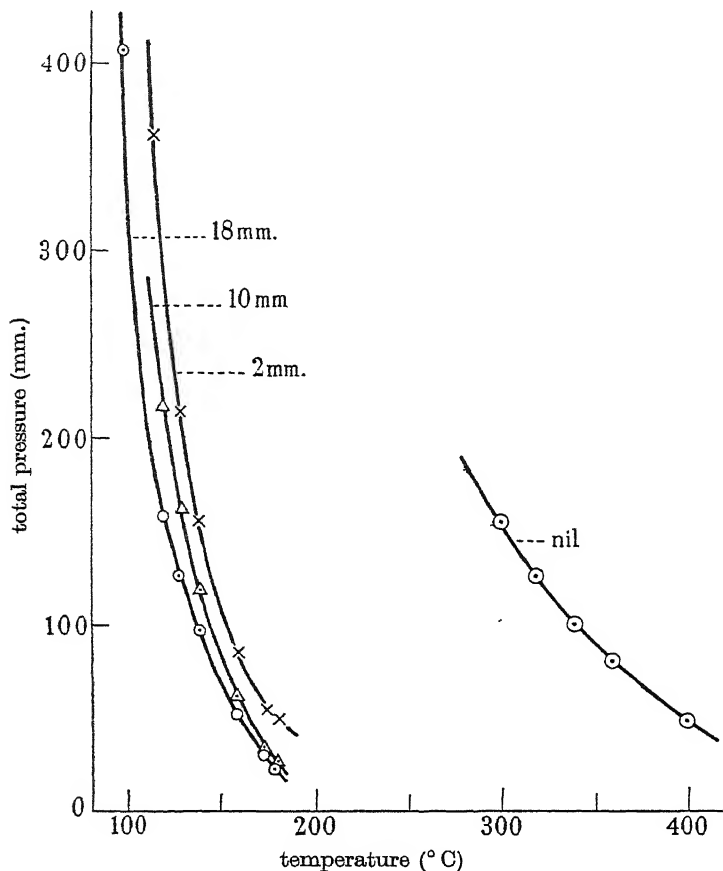


FIGURE 2. Variation of minimum ignition pressure ($H_2 + Cl_2$) with temperature for hydrogen and chlorine, with 0, 2, 10, and 18 mm. of chloropicrin. Pyrex vessel *A*.

When the gases flow in through the normal entry tube at temperatures above 340° C, the boundary remains low. At pressures just inside this boundary, the ignition is accompanied by a click, clearly audible but not as loud as at lower temperatures. The flame is not nearly so bright as at lower temperatures, and lasts an appreciable fraction of a second. When the gases flow in through the heated entry tube, the ignition boundary is much higher than for the normal entry tube. In addition, when the flame appears, it is very feeble and lasts several seconds, and could well be described as resembling the glow of phosphorus, except that its colour is a little more yellow.

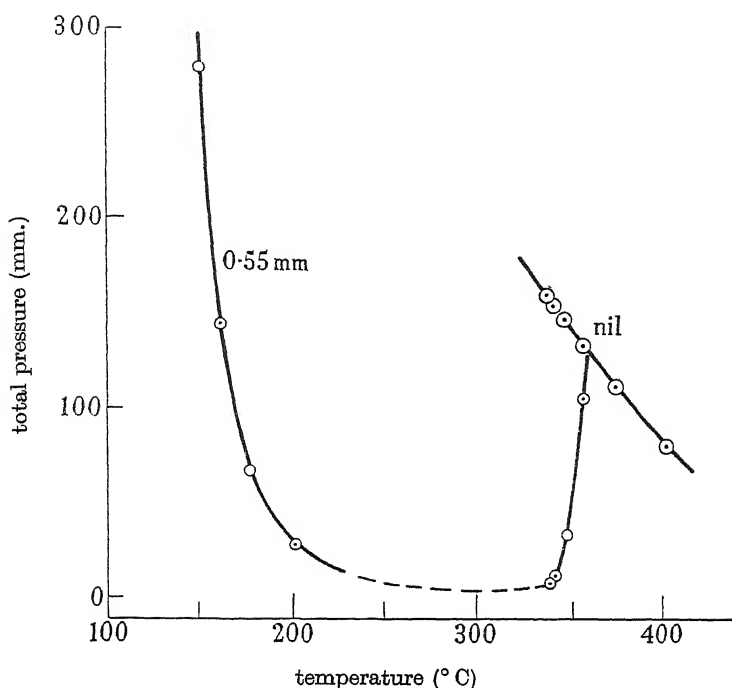


FIGURE 3. Variation of minimum ignition pressure ($H_2 + Cl_2$) with temperature for hydrogen and chlorine, with 0 and 0.55 mm. of chloropicrin. Pyrex vessel B.

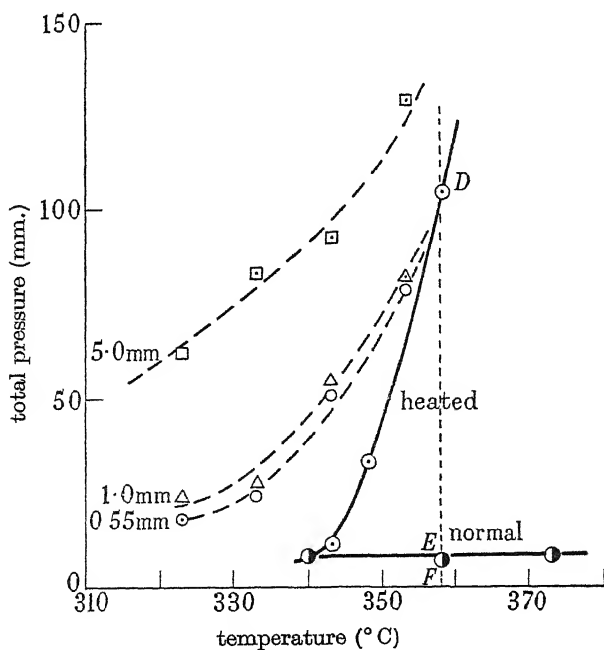


FIGURE 4. The effect of mode of entry on the upper temperature limit to the region of sensitized ignition in vessel B. Broken lines show effect of different pressures of chloropicrin with a capillary inlet. Full lines show effect of 'normal' (◐) and 'heated' (◉) entry tubes, with 0.55 mm. of chloropicrin. See figure 9 for reaction rates along DEF.

The 'backfire' phenomenon

An interesting phenomenon occurs above 350°C and at pressures just below the boundary for sensitized mixtures flowing in through the heated entry tube. As has been said, there is no flame visible in the reaction vessel. There is, however, a bright flame which starts in the hot entry tube and travels back into the cold mixing vessel, accompanied by a very loud click and a considerable pressure pulse. This phenomenon was repeatedly confirmed by several observers, who separately watched the reaction vessel, observation points cleared on the enamelled connecting tubing, and the mixing vessel. Its implications are discussed later. When the subsidiary furnace on the normal entry tube is kept at the temperature of the main

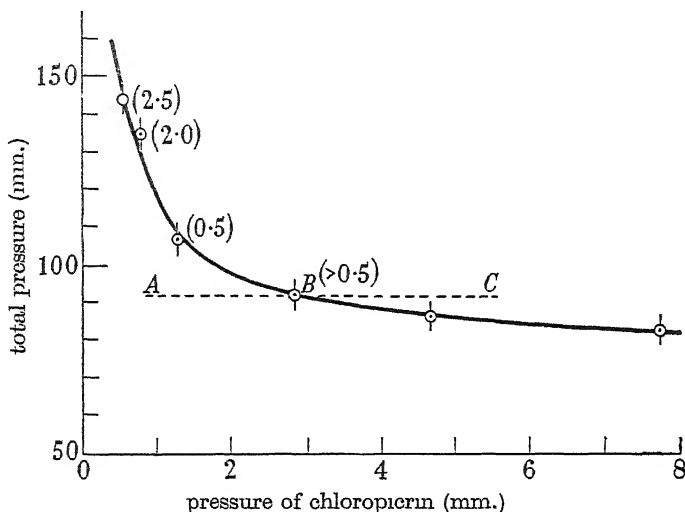


FIGURE 5. The variation of the minimum ignition pressure ($\text{H}_2 + \text{Cl}_2$) with chloropicrin pressure at 162°C in reaction vessel *B* (heated entry). Introduction periods (seconds) in brackets. See figures 7 and 8 for reaction rates along *ABC*.

furnace, the gases admitted through the normal entry tube behave in precisely the same manner as when flowing through the heated entry tube. The ignition boundary then rises steeply with increase of temperature, and the 'backfire' phenomenon described above appears just below this new position of the boundary. It seems clear from this that the different effects observed for the normal (unheated) entry tube and heated entry tube are due solely to the length of narrow tube which is at the furnace temperature.

Variation of the ignition boundary with chloropicrin pressure at 162°C

Figure 5 shows the variation at 162°C of the boundary of ignition of equimolar mixtures in vessel *B* (heated entry tube) as the chloropicrin pressure is increased. The equation relating the ignition pressure P and the chloropicrin concentration C_p is clearly some form of hyperbola with asymptotes other than $P=0$ and $C_p=0$. At the lower chloropicrin concentrations, there is a short induction period, and representative values in seconds are given on the curve.

Variation of the ignition boundary with proportions of reactants

In this same temperature range (100 to 200° C) the position of the boundary varies with the proportion of chlorine in the mixture. The variation of the boundary with the percentage of chlorine in the mixture, in the presence of a fixed concentration (0.55 mm.) of chloropicrin, at a temperature of 177° C, is shown in figure 6. The ignition pressure is at a minimum at about 65 to 70 % of chlorine.

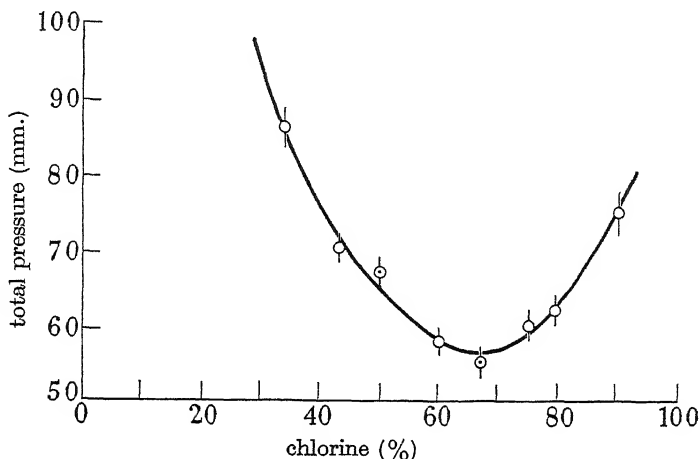


FIGURE 6. The variation of the minimum ignition pressure with the percentage of chlorine in the mixture at 177° C in reaction vessel B (heated entry) in the presence of 0.55 mm. of chloropicrin.

The percentage of reaction completed after ignition and during slow reaction

The question arises, however, what these 'visible flame' boundaries for the sensitized mixtures mean in terms of the amount of reaction completed. This was investigated near the low temperature boundary, at 162° C, by determining the percentage conversion to hydrogen chloride of a given mixture of hydrogen and chlorine, with various concentrations of chloropicrin added, at a given time. The initial compositions of the mixtures thus lie on the line *ABC* passing through the ignition boundary of figure 5. In figure 7 the percentage conversion, at 5 sec. after admission, is plotted against the pressure of chloropicrin, for 93 mm. of $H_2 + Cl_2$. The position of the sharp variation in percentage conversion coincides with the point *B* on the visible flame boundary of figure 5. In order to confirm that the region of very low percentage completion of reaction in figure 7 is not due to an induction period before an accelerating reaction, the percentage conversion of suitable mixtures was determined at intervals up to 40 sec., with the results plotted in figure 8. It will be seen that the percentage conversion at a given time is practically independent of the amount of chloropicrin used, and that it does not reach large values even after half a minute. In the absence of chloropicrin, the amount of conversion after 40 sec. was less than 1 %.

The variation of the extent of reaction as the higher temperature boundaries are approached was investigated at 358° C, using a fixed amount of chloropicrin and varying the total pressure of reactants, so moving along the line *DEF* of figure 4.

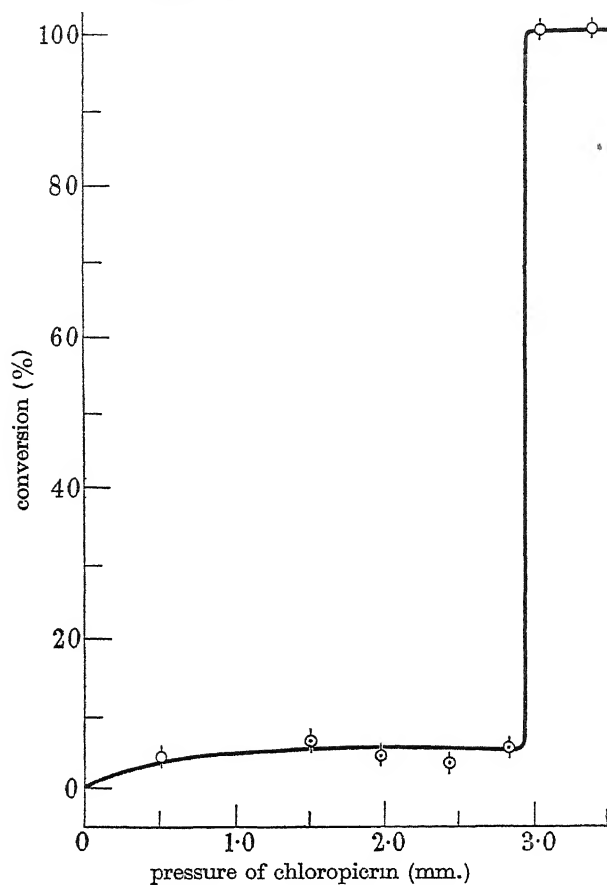


FIGURE 7. The dependence of the percentage conversion of 93 mm. of hydrogen and chlorine in 5 sec. at 162°C , upon the pressure of chloropicrin.

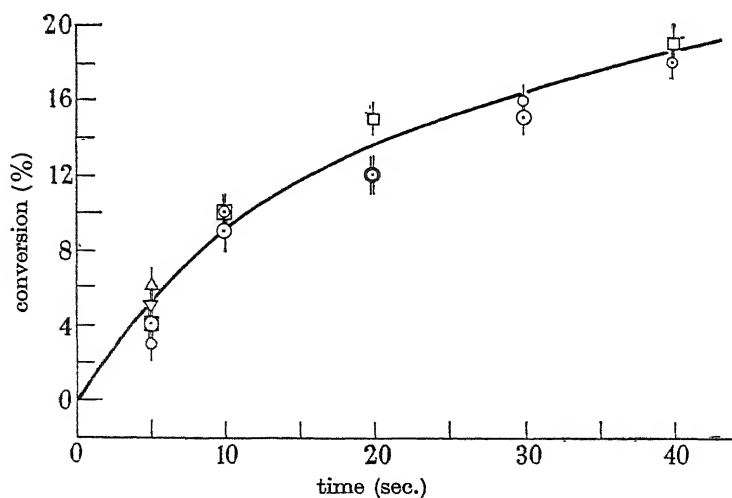


FIGURE 8. The percentage conversion of 93 mm. of hydrogen and chlorine at different time intervals. 162°C . Concentration of chloropicrin: ○ 0.55 mm.; △ 1.55 mm.; □ 1.98 mm.; ○ 2.39 mm.; ▽ 2.71 mm.

It is plain from figure 9 that the percentage conversion after $2 (\pm 0.25)$ sec. varies markedly with the method of entry. The percentage conversion rises much more rapidly with increasing total pressure for the normal inlet tube. With admission through the normal tube, the flame first appears in the reaction vessel with about 8 mm. of gases giving about 50% conversion in two seconds, whereas the same

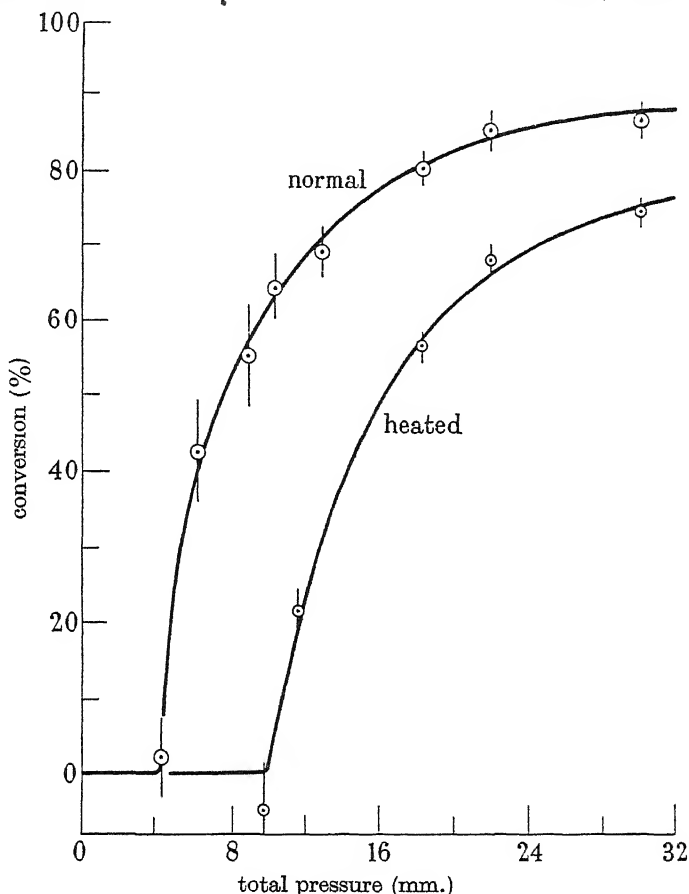


FIGURE 9. The percentage conversion of ($\text{H}_2 + \text{Cl}_2$) at various total pressures in $2 \pm \frac{1}{4}$ sec. at 358°C , in the presence of 0.55 mm. of chloropicrin, using the 'normal' and 'heated' entry tubes of vessel B.

pressure of 8 mm. admitted through the heated entry tube gives no flame and less than 5% conversion in two seconds. With the heated entry tube, the percentage conversion after two seconds has risen to more than 80%, and probably much higher, before the flame appears. It is clear that the higher temperature boundary is not defined by the same clear demarcation between very little reaction and complete reaction which characterizes the lower limit.

Effect of the decomposition products

The effect of the decomposition products of chloropicrin, nitrosyl chloride and carbonyl chloride, upon the ignition boundary of equimolar mixtures of hydrogen and chlorine was determined. The ignition limits in the presence of 0.1 mm. of

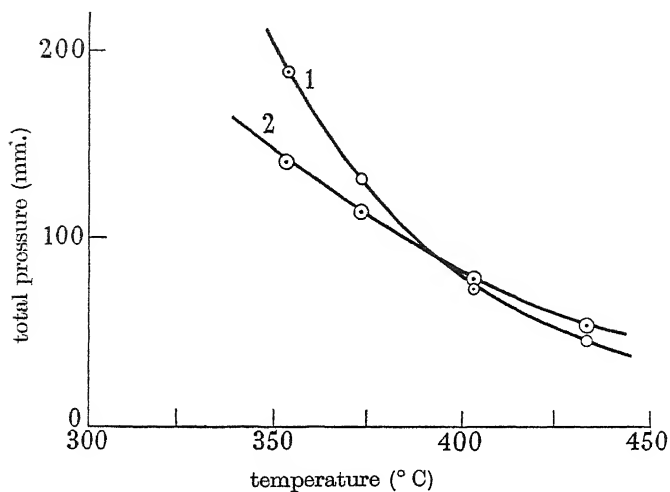


FIGURE 10. The minimum ignition pressure ($\text{H}_2 + \text{Cl}_2$) in the presence of 0.1 mm. of nitrosyl chloride (curve 1) and no nitrosyl chloride (curve 2). Pyrex vessel *B*.

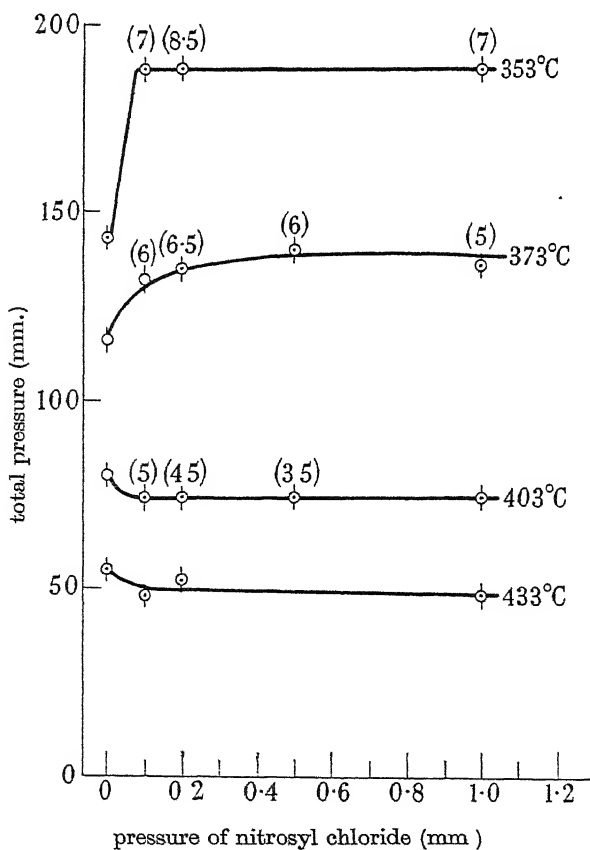


FIGURE 11. The effect on the minimum ignition pressure at different temperatures of adding nitrosyl chloride. Induction periods in brackets. Heated entry tube of vessel *B*.

nitrosyl chloride are compared with the limits for the pure gases in figure 10, and the effect of varying the nitrosyl chloride concentration is shown in figure 11. It will be seen that at temperatures above 390°C , nitrosyl chloride has a slight sensitizing effect. At lower temperatures, the ignition pressure rises steeply (to a practically constant value) as the pressure of nitrosyl chloride is increased. There is a short induction period, representative values being shown on the curves of figure 11 for each temperature. These values include the time of flow into the vessel, which was less than 0.5 sec., the short capillary was present for these experiments. Carbonyl chloride appears to have little effect on the ignition boundary. The position of the limits, with or without nitrosyl chloride, is not affected by differences in the temperature of the entry tubes, the same limit being observed for heated and normal entry tubes.

GENERAL DISCUSSION OF RESULTS

The results presented show that chloropicrin causes a considerable extension of the ignition region of mixtures of hydrogen and chlorine to comparatively low temperatures and pressures. The region of sensitized ignitions, which in general shape is not unlike the cold flame region of some ethers and hydrocarbons, has a clear and reproducible boundary on the low temperature-pressure side, marking a sharp change from slow reaction to ignition as the pressure or temperature is increased. The position of this lower boundary is lowered by an increase in the concentration of chloropicrin added, but appears to be independent of the way a particular mixture of gases flows into the vessel. At higher temperatures (above 200°C) the boundary falls at first to very low pressures, and in the absence of any profound alteration in the composition of the mixture entering the vessel, the boundary might be expected to remain at very low pressures, until it merges into the ignition boundary of the unsensitized mixtures. However, it is found that with some conditions of gas flow into the vessel, the boundary of visible ignition rises steeply at temperatures above 320°C . The main condition which favours this rise appears to be the presence of a long heated portion of the entry tube of narrow diameter. When a short portion of the entry tube is heated, with the same total path length, the boundary is much lower. It has been shown in part VIII that estimates can be made of the proportion of chloropicrin decomposed within the entry tube, by treating the admission tap as a nozzle and hence calculating the time of flow of gases into the vessel. Figure 18 (part VIII) shows that above 320°C a substantial proportion of the chloropicrin entering the vessel will be decomposed within the time spent in even a short entry tube. The curve of figure 18 (part VIII) lies in the same temperature region as the curves of figure 4 of this paper. Consequently, the position of the boundaries of figure 4 might be explained if it is postulated that the onset of ignition depends upon the presence of a certain amount of undecomposed chloropicrin which can initiate the hydrogen-chlorine chains. This amount is exceeded in the vessel after entry through the short heated tube and ignition occurs; during entry through the long heated tube, ignition is prevented inside the entry tube, in spite of a suitable concentration of chloropicrin,

by the efficient chain termination on the walls, and by the time the gases reach the vessel, the amount of chloropierin is too low to cause ignition

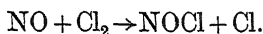
There are, however, two significant observations which suggest that this simple explanation of the rising boundary above 320°C must be modified. The first is the very remarkable 'backfire' phenomenon observed when using the entry tube with the longer heated portion, at pressures just below each boundary of visible ignition in the region of 360°C . Here, a flame propagates back into the cold mixing vessel, and can be seen at the beginning of the entry tube and in the connecting tube (also of small diameter), but no flame is seen in the hot reaction vessel nor in the adjacent portions of the entry tube. It is evident that, under these conditions, ignition is not suppressed solely by the walls of the narrow tubes. Some other substance, capable of terminating chains, must be appearing towards the end of the entry tube, and as such a pronounced effect can hardly be due to hydrogen chloride (because less HCl is formed in a given time for entry through the heated tube, than for entry through the normal tube, figure 9), it is reasonable to ascribe it to the substances formed by decomposition of the chloropierin during its flow through the longer heated entry tube.

The second observation is that in this region of rising ignition boundary, increase of chloropierin concentration markedly raises the boundary, figure 4. If the minimum pressure of reactants necessary for ignition is solely dependent on the amount of undecomposed chloropierin arriving in the vessel, the boundary would be expected to fall with increase of chloropierin concentration. The fact that it rises cannot be due to an increased production of HCl during the time of entry, for this would necessitate an increased rate of reaction, and it can hardly be postulated that an increased rate can make ignition less likely. Once again, it is clear that a product of the chloropierin decomposition is exerting a considerable inhibiting influence on some stage of the reaction leading to ignition.

The ignition of mixtures of hydrogen, chlorine, and chloropierin, whose temperature and concentration lie well within the region of sensitized ignition, can be suppressed by adding nitrosyl chloride to the mixture before admission. For example, at 273°C , a total pressure of 86 mm. of $\text{H}_2 + \text{Cl}_2$ with 3.2 mm. of chloropierin gives a bright hot flame, and from figures 3 and 4 is clearly well within the ignition region. If 0.50 mm. of nitrosyl chloride are added, the flame is suppressed, and analysis shows that the amount of reaction which has occurred in 10 sec. is less than $5 \pm 3\%$.

The effect of nitrosyl chloride itself upon the ignition boundary of 1:1 mixtures of hydrogen and chlorine is interesting. Dainton (1947) showed that very small quantities of nitrosyl chloride strongly inhibit the photochemical reaction between carbon monoxide and chlorine at room temperature. The effect was ascribed to the shortening of the reaction chains by the very efficient chain terminating step $\text{Cl} + \text{NOCl} \rightarrow \text{NO} + \text{Cl}_2 + 19.5 \text{ kcal. } (k_3)$. This reaction appears to have a low energy of activation (Ashmore & Norrish 1950*b*, part X) and to be fast at room temperatures. It would be less important at temperatures of 350 to 450°C , relative to atom-molecule reactions of greater activation energies, but it would be expected to shorten the chain length of the hydrogen-chlorine chains in the

presence of nitrosyl chloride, thereby raising the ignition boundary. It was therefore rather surprising to find that between 380 and 480° C the ignition boundaries of the two systems cross (figure 10); below 400° C, there is a considerable raising of the ignition boundary on the addition of 0.1 mm. of nitrosyl chloride, but above 400° C the same addition produces a lowering of the ignition boundary. Moreover, the position of the boundary rapidly becomes independent of the nitrosyl chloride concentration as this is increased, figure 11. This observation might suggest that the nitrosyl chloride suppresses the chain reaction almost completely, leaving perhaps a bimolecular reaction between hydrogen and chlorine which can give a thermal ignition at higher pressures. It seems impossible, however, that the curve of such ignition pressures against temperature can actually cross the corresponding curve in the absence of nitrosyl chloride, when both bimolecular and chain reaction should contribute to the rate. When it is recalled that nitrosyl chloride sensitizes the ignition of hydrogen-oxygen mixtures by providing, in this temperature region, initiation and branching steps in the ignition chains (Dainton & Norrish 1941), it clearly becomes desirable to investigate whether the effect in the hydrogen-chlorine-nitrosyl chloride system can be explained by initiation or branching steps involving nitrosyl chloride molecules. Another possibility is that the nitric oxide formed by decomposition of the nitrosyl chloride can restore the chlorine atom concentration by the reaction



This reaction, the reverse of k_3 , has an activation energy of about 20 kcal. and will only be important at high temperatures. Whatever the detailed kinetic explanation of the effect of nitrosyl chloride, it remains true that below 400° C the resultant effect is one of inhibition, and that the nitrosyl chloride formed by the decomposition of the chloropierin, in the chloropierin sensitized reaction, will have the opposite effect to the chloropierin itself. On the other hand, since carbonyl chloride has been shown experimentally to have little effect on the limit there is no need to postulate any effects due to the carbonyl chloride formed in parallel with the nitrosyl chloride in the reaction sensitized by chloropierin.

CONCLUSIONS

The general arguments outlined above suggest that in the temperatures range of 100 to 200° C, the decomposition of the chloropierin provides a large number of centres which can start the hydrogen-chlorine chains. At temperatures above 200° C, the more rapid decomposition of the chloropierin gives sufficient nitrosyl chloride to introduce powerful chain-terminating reactions, and according as the decomposition is partial or complete by the time the reactants reach the reaction vessel, the ignition is partially or completely suppressed unless the concentration of the reactants is high. These conclusions are in good agreement with those drawn from the work described in part VIII. They will be examined in more detail in part X in relation to a kinetic scheme for the decomposition of chloropierin.

The authors wish to thank the Governing Body of Emmanuel College, Cambridge, for a Research Studentship given in 1938 to one of us (P.G.A.) and for a Research Fellowship (P.G.A.) which enabled this work to be resumed in 1946. They also wish to thank the Department of Scientific and Industrial Research for a Research Maintenance Grant in 1938 (P.G.A.)

REFERENCES

- Ashmore, P. G. & Norrish, R. G. W. 1950a *Proc. Roy. Soc. A*, **203**, 454. (part VIII).
 Ashmore, P. G. & Norrish, R. G. W. 1950b *Proc. Roy. Soc. A*, **204** (in the Press) (part X).
 Christiansen, J. A. 1929 *Z. phys. Chem. B*, **2**, 405.
 Dainton, F. S. 1947 *Trans. Faraday Soc.* **43**, 365.
 Dainton, F. S. & Norrish, R. G. W. 1941 *Proc. Roy. Soc. A*, **177**, 393, 411, 421.
 Khodschanian, S. & Kornfeld, A. 1937 *Z. Phys. Chem. B*, **35**, 403.
 Morris, J. C. & Pease, R. N. 1939 *J. Amer. Chem. Soc.* **61**, 391.
 Pease, R. N. 1934 *J. Amer. Chem. Soc.* **56**, 2388.
 Sagulin, A. B. 1928 *Z. Phys. Chem. B*, **1**, 275.
 Semenov, N. 1935 *Chemical kinetics and chain reactions*, p. 115. Oxford University Press.

The kinetics of the thermal decomposition of normal paraffin hydrocarbons

III. Activation energies and possible mechanisms of molecular reactions

BY K. U. INGOLD, F. J. STUBBS AND SIR CYRIL HINSHELWOOD, F.R.S.

Physical Chemistry Laboratory, University of Oxford

(Received 14 June 1950)

In part I it was concluded that the nitric oxide-inhibited decomposition of paraffins probably represents a molecular reaction. Further experiments in which the presence of hydrogen causes a marked increase in the normal reaction but not of the inhibited reaction strengthen this conclusion, by diminishing still further the likelihood that the inhibited reaction is a chain process not suppressible by nitric oxide.

Experiments on variation of the surface/volume ratio and on the coating of the vessel surface with potassium chloride have been made for the normal reaction and for the reaction inhibited by nitric oxide and by propylene respectively. The effect of the surface change is either negligible or, in certain cases, to accelerate a condensation reaction which may vitiate the measurement of the true decomposition rate.

Over limited ranges the rate of reaction, r_{∞} , is connected with the pressure by the relation

$$r_{\infty} = Ap_0 + Bp_0^2,$$

but this is probably an approximation for an expression of the form

$$r_{\infty} = \frac{ap_0^2}{1 + a'p_0} + \frac{bp_0^2}{1 + b'p_0},$$

the reaction mechanism being composite. A reaction nearly of the first order predominates at lower pressures and one nearly of the second order at higher pressures.

The activation energy of the nitric oxide-inhibited reaction with *n*-pentane and *n*-heptane increases steeply as the pressure falls, and by suitable extrapolations the values corresponding to the two assumed components can be estimated. With ethane and propane the activation energy varies very little with pressure while *n*-butane is an intermediate case.

The different modes of reaction appear not to be connected with the different possible positions of rupture of the molecule. Their explanation probably demands an extension of the usual theory of unimolecular reactions. In a complex molecule a large amount of energy distributed in many degrees of freedom eventually causes decomposition by the interference of normal vibration modes, while alternatively a much smaller amount can cause reaction if it is so communicated that the critical bond breakages can occur before the energy has been shared throughout the molecule.

1. INTRODUCTION

Parts I and II of this series (Stubbs & Hinshelwood 1950*a, b*) have provided *prima facie* evidence that the nitric oxide-inhibited decomposition of the normal paraffin hydrocarbons is apparently a molecular rearrangement process, persisting after the radical-induced chains have been suppressed. The variation of reaction rate with pressure was found in general to be satisfied over a considerable range by an equation of the form $r_{\infty} = Ap_0 + Bp_0^2$. The reaction order thus increases with pressure, being almost the first at lower pressures and tending towards the second at higher. This pressure dependence is not consistent with the uniform transition from second order to first predicted for a single unimolecular reaction dependent upon activation by collision. Various possibilities were examined to explain why the reaction was not of the normal unimolecular type, and it was decided to seek further information by making a detailed study of the variation with pressure of the activation energies for a series of normal paraffins.

The principal result of the investigation is to show that whereas the activation energy for the fully inhibited decomposition is nearly independent of the initial hydrocarbon pressure with ethane and propane, with *n*-pentane and *n*-heptane—and presumably with other higher paraffins—it decreases markedly with increasing initial pressure. That for the decomposition of *n*-butane at first decreases as the pressure increases but becomes constant above 200 mm. The implications of this will be considered in a later section.

2. SYSTEMATIC STUDY OF THE ACTIVATION ENERGY-PRESSURE RELATIONS

The experimental technique has been described in part I. At constant temperature the rate of the fully inhibited decomposition (r_{∞}) was determined for a series of initial hydrocarbon pressures, this procedure was repeated for 10° intervals over a 60° C range. Rate-pressure curves of the form mentioned above were fitted to the experimental points, and from these curves the rate at a particular pressure could be read off. In this way it was possible to determine the activation energy for a number of different pressures for each paraffin.

The rates were referred to a constant concentration in each case rather than to a constant pressure, e.g. at 550° C the rate would be read off at a pressure of 102.4 mm., since this pressure reduces to 100 mm. at 530° C (which was taken as the standard temperature).

The pentane and heptane used were very pure specimens which were gifts from the Anglo-Iranian Oil Company and from Imperial Chemical Industries. The lower hydrocarbons were 'pure' commercial specimens subjected to distillation before use.

Ethane. The fully inhibited decomposition was studied over a pressure range from 50 to 600 mm. The variation of rate with nitric oxide concentration was investigated at 600, 620 and 640° C, the curve for 600° C agreeing closely with the results of Hobbs & Hinshelwood (1938). The amount of nitric oxide necessary for maximum inhibition was greater than for the other hydrocarbons, and also increased with increasing pressure as pointed out by Partington, Stubbs & Hinshelwood (1949).

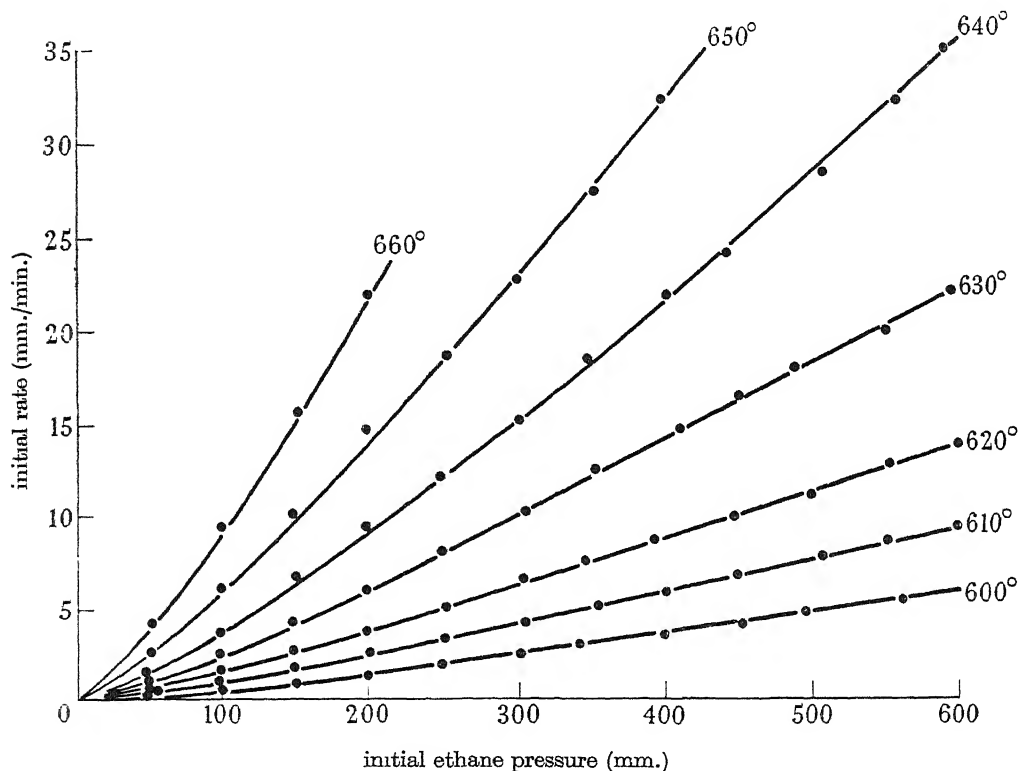


FIGURE 1. Fully inhibited decomposition of ethane.

The plot of initial rate against ethane pressure for the fully inhibited decomposition at various temperatures is shown in figure 1. The rate-pressure relationship is linear from 250 to 600 mm., but markedly curved below 250 mm. The behaviour predicted for a single unimolecular reaction with activation by collision is a transformation from first order to second order at low pressures. The rate would be given by $r_{\infty} = k_1 p_0^2 / (1 + k_2 / k_3 p_0)$, where k_1 and k_2 are constants and k_3 may be a constant or vary with pressure. For the simplest case where k_3 is constant

$$p_0 / r_{\infty} = 1 / k_1 p_0 + k_2 / k_1 k_3.$$

This equation indicates that the plot of p_0 / r_{∞} against $1 / p_0$ should be linear, and this is in fact found to be so for ethane, though for none of the other paraffins.

A few experiments were made to compare the present results with those of other workers. The Δp -time curve for the decomposition of 330 mm. of ethane at 607°C, without nitric oxide, was identical with a similar one based on the results of Steacie & Shane (1940). These workers state, however, that 1 % of nitric oxide is sufficient for maximum inhibition, while the present investigation shows that at this temperature the amount of nitric oxide necessary for really complete inhibition is nearer 10 %. An experiment with 330 mm. of ethane and 3.3 mm. of nitric oxide gave a greater initial rate and larger Δp values than those observed by Steacie & Shane, whereas with 33 mm. of nitric oxide the Δp -time curve was nearly coincident with theirs.

The apparent chain lengths at a pressure of 200 mm., and at temperatures of 600, 620 and 640°C, were 4.3, 3.7 and 3.4 respectively. These values are in close agreement with those obtained by Steacie & Shane, but smaller than those of Hobbs & Hinshelwood (1938) and Staveley (1937).

Propane. The fully inhibited decomposition was studied for pressures from 50 to 500 mm. At lower pressures the rate-pressure relations are given by an equation of the type

$$r_{\infty} = Ap_0 + Bp_0^2, \quad (1)$$

but at higher pressures, and particularly at the higher temperatures, the experimental curve becomes nearly linear. The overall rate-pressure relationship is exactly satisfied by an equation of the form

$$r_{\infty} = Ap_0 + \frac{bp_0^2}{1 + b'p_0}. \quad (2)$$

At low pressures this approximates in form to equation (1), but at higher pressures approximates to the equation of a straight line. The values of the constants A , b and b' for the various temperatures are given in table 1.

TABLE 1. PROPANE: VALUES OF CONSTANTS IN THE EQUATION

$$r_{\infty} = Ap_0 + \frac{bp_0^2}{1 + b'p_0}$$

temp. (°C)	$A \times 10^2$	$b \times 10^4$	$b' \times 10^2$
550	0.42	0.16	0.16
560	0.67	0.27	0.16
570	1.1	0.48	0.20
580	1.7	0.80	0.21
590	2.5	1.46	0.23
600	4.0	2.36	0.25
610	6.0	4.0	0.27

n-Butane. The fully inhibited decomposition was studied from 50 to 600 mm. As with propane, but less markedly, at high pressures the rate-pressure relationship tends to become linear. For all practical purposes, however, equation (1) fits closely enough over the pressure range investigated. The values of the constants A and B at different temperatures are given in table 2.

n-Pentane. This was studied at pressures from 50 to 450 mm. The rate-pressure relationships conform closely to equation (1), and the values of the constants A and

B are given in table 2. The results for 530 and 550°C were identical with those obtained in the previous work (Stubbs & Hinshelwood 1950*b*). Figure 2 shows a plot of $\log r_\infty$ against the reciprocal of the absolute temperature for various pressures of *n*-pentane. For each pressure the points lie very close to a straight line. This was equally true for the other hydrocarbons.

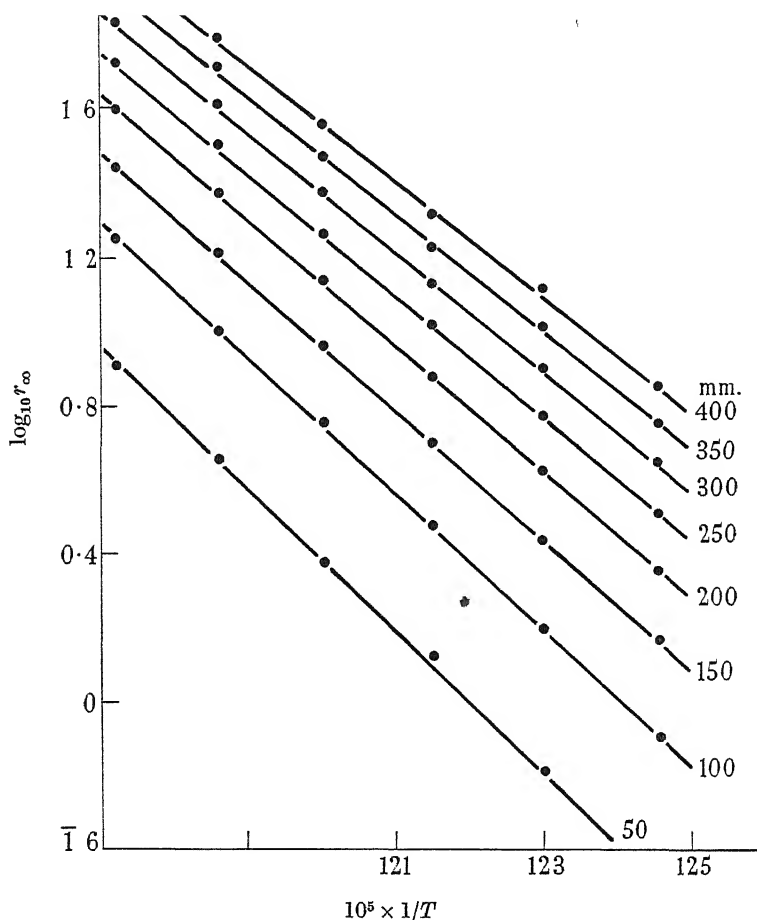


FIGURE 2. Plot of $\log_{10} r_\infty$ against $1/T$ for *n*-pentane.

n-Heptane. This was studied at pressures from 25 to 275 mm. The rate-pressure relationship conforms closely to equation (1), and the values of the constants A and B are given in table 2

The variation with pressure of the activation energy of the fully inhibited decomposition of the different hydrocarbons is summarized in table 3. For comparative purposes the results are plotted in figure 3. Some interesting conclusions can be drawn from these results, but the theoretical discussion must be preceded by a consideration of various matters which concern the significance of the measurements themselves (§§ 3, 4, 5 and 6).

TABLE 2. CONSTANTS A AND B

temp. ($^{\circ}$ C)	<i>n</i> -butane		<i>n</i> -pentane		<i>n</i> -heptane	
	$A \times 10^2$	$B \times 10^5$	$A \times 10^2$	$B \times 10^5$	$A \times 10^2$	$B \times 10^5$
510	—	—	—	—	0.58	4.2
520	—	—	—	—	1.2	6.0
530	0.55	2.1	0.47	3.12	2.1	9.5
540	0.96	2.9	0.90	5.4	4.0	13.5
550	1.6	4.5	2.2	6.5	7.8	16.0
560	2.5	6.7	4.2	10.2	13.3	22.5
570	4.0	10.0	7.5	16.2	—	—
580	6.2	14.0	13.2	23.0	—	—

TABLE 3. ACTIVATION ENERGY AND PRESSURE

initial pressure (mm.)	activation energy (kcal./g. mol.)				
	ethane	propane	<i>n</i> -butane	<i>n</i> -pentane	<i>n</i> -heptane
50	76.4	68.2	70.8	88.8	79.7
100	74.7	68.7	64.9	85.1	73.6
150	74.7	69.1	61.3	81.0	70.5
200	74.7	70.1	61.2	78.3	68.7
250	74.7	71.1	60.6	76.0	67.2
300	74.7	71.2	60.7	74.2	—
350	74.7	70.8	60.8	73.0	—
400	74.7	71.0	60.9	71.9	—
450	74.7	70.2	60.6	—	—
500	74.7	—	60.6	—	—
550	74.7	—	59.9	—	—

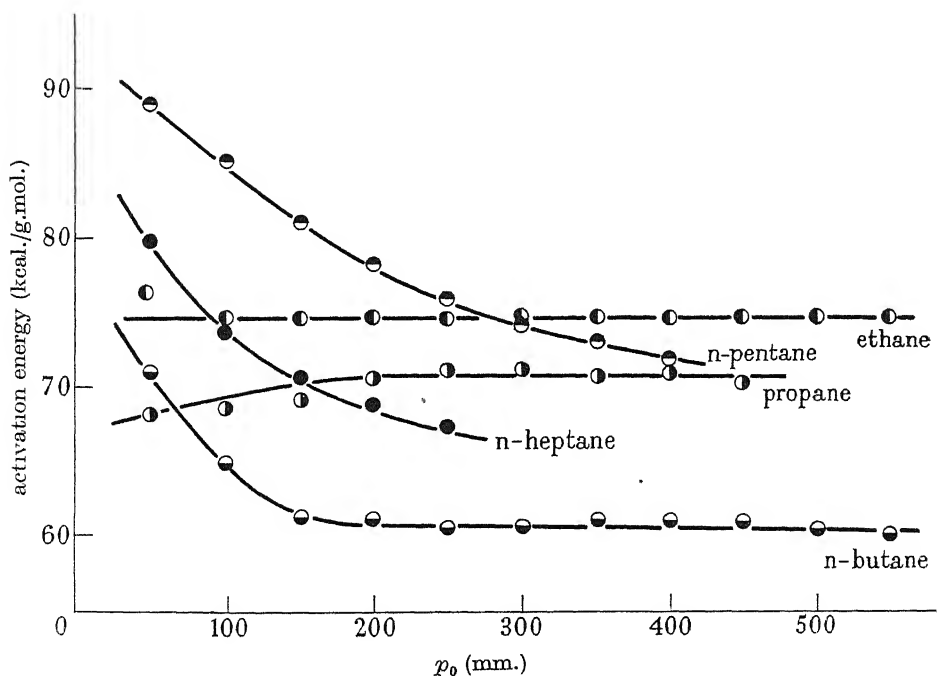


FIGURE 3. Variation of activation energy with pressure.

3. SIGNIFICANCE OF THE EXPERIMENTAL ACTIVATION ENERGIES

For the higher hydrocarbons the activation energy falls off rapidly with increasing pressure. At first sight, and without paying much attention to the theoretical aspect, one would conclude that a process with a higher activation energy predominates at low pressures and one with a lower activation energy at high pressures. If this is so, the measured value is really a function of two different activation energies, and in principle the logarithm of the total rate plotted against $1/T$ should no longer be a straight line. Numerical trial, however, with two activation energies of 85 and 65 kcal. and with the two rates assumed to be equal at 530° C showed that a plot of the logarithm of the sum of the individual rates had negligible curvature over a temperature range of 100° C. In all cases the plot of $\log r_{\infty}$ against $1/T$ was in fact a straight line within the limit of experimental error.

The value of E determined from the slope of this line has in itself no physical significance for pressures in the intermediate range. That obtained by extrapolation to zero pressure will, however, be the correct activation energy of the one component of the composite reaction mechanism, and similarly the limiting value reached as the pressure increases will be the activation energy of the other component. The limiting value is very readily obtained, and the extrapolation to zero pressure is also reasonably good. The various extrapolated results are summarized in table 5 (§7)

4. EFFECT OF SURFACE

The results found in the present investigation with *n*-pentane and *n*-heptane were identical with those obtained previously (part II), but with propane and *n*-butane the rates obtained for the fully inhibited decomposition were somewhat greater than before. The butane samples used on the two occasions were of different origin, but direct comparison showed their behaviour to be now identical, and infra-red analysis (carried out by Mr R. L. Williams and Dr H. W. Thompson, F.R.S.) showed them both to be at least 99.5 % pure. The induction period, which in part I was ascribed to a disturbing transitory surface reaction attended by a pressure decrease, had now almost disappeared. It seemed likely, therefore, that the earlier results were in fact a little too low in the case of the propane and butane because this initial disturbance had not been fully allowed for. In between the two sets of experiments the state of the surface had presumably changed.

This fact suggested the need for a fuller investigation of the surface effects, and a series of experiments was therefore carried out with a tube-packed silica vessel of volume 250 ml. having a surface/volume ratio 5.2 times as great as the original silica vessel (which will be referred to as the standard vessel).

In the packed vessel the induction period was in evidence for the fully inhibited decomposition of all the paraffins. With 200 mm. butane and 15 mm. nitric oxide at 530° C, for example, it was 2.2 min., while in the standard vessel it had varied from 1.5 to 0.3 min.

With *n*-heptane and with *n*-pentane the complete inhibition curves at 530° C were coincident for the packed and the standard vessels. The results for heptane are shown in figure 4.

With *n*-butane the rate of the normal decomposition was unaffected by the increase in surface, but that of the fully inhibited decomposition was lowered appreciably. Rather similar results were found for propane and for ethane. The ratio of rates in various circumstances is given in table 4.

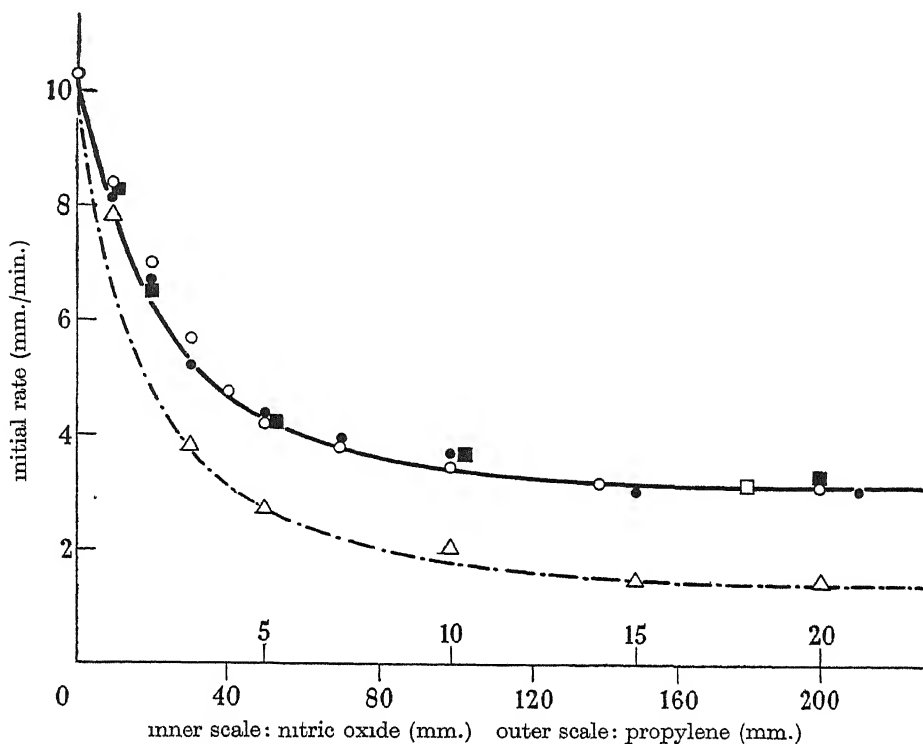


FIGURE 4 Thermal decomposition of 100 mm. *n*-heptane at 530° C. Nitric oxide as inhibitor: ○ standard vessel; ● packed vessel; △ KCl coated vessel. Propylene as inhibitor: □ standard vessel; ■ KCl coated vessel.

It may be concluded that the uncontrolled variations occasionally appearing with propane and butane are of the same nature as those caused by the increase of surface. In the light of the conclusions of part I the effect of the increased surface, where it is observed, is to promote a certain secondary reaction in presence of nitric oxide, with an attendant decrease of pressure which causes an apparent lowering of the decomposition rate.

The effect of varying the nature of the surface by a coating of potassium chloride was next studied. In the case of propane, *n*-butane and *n*-heptane, the normal reaction is unaffected, but the rate of the fully inhibited reaction is considerably reduced. The results for heptane, which are typical, are shown in figure 4.

At first sight this might seem to indicate that the fully inhibited reaction itself involves chains which the potassium chloride surface can break. This interpretation would, however, be incorrect. In the standard vessel propylene reduces the rate to exactly the same limit as nitric oxide. Further experiments now show that in the potassium chloride-coated vessel propylene gives the same limit as in the standard

vessel. This is shown in figure 4, where the results for propylene are plotted with a simple change of horizontal scale. With propylene as inhibitor there is, moreover, no induction period.

The apparent reduction in rate caused by increased surface and by the potassium chloride-coated surface is therefore a disturbance caused by some interaction in the system, surface, nitric oxide, paraffin. This leads to a pressure decrease which vitiates the measurement of the true decomposition rate.

TABLE 4. EFFECT OF SURFACE

hydrocarbon	temp. (° C)	initial pressure (mm.)	standard vessel	packed vessel	ratio rate in standard vessel rate in packed vessel
ethane	600	200	1.3	1.1	1.18
	630	200	5.8	5.4	1.07
	650	200	15.0	13.6	1.03
propane	530	100	0.2	0.1	2.0
	570	100	1.75	1.02	1.72
	600	100	5.6	4.5	1.25
	560	51	0.40	0.24	1.70
	560	96	0.85	0.51	1.67
	560	146	1.42	0.90	1.60
	560	207	2.21	1.41	1.67
	560	278	3.20	2.01	1.60
	560	339	4.25	2.70	1.57
	560	414	5.57	3.60	1.55
<i>n</i> -butane	560	100	2.8	2.0	1.40
	570	100	4.7	4.0	1.17
	530	200	2.1	1.53	1.37
	570	200	11.8	9.5	1.24
<i>n</i> -pentane	530	100	0.70	0.75	1.0
<i>n</i> -heptane	530	100	3.1	3.0	1.0

In no case has the surface any effect on the uninhibited reaction. This indicates that in the normal decomposition chains are broken in the gas phase and not on the surface. The effect of the potassium chloride surface in reducing the limiting rate obtained with nitric oxide is thus unlikely in any case to be attributable to some extra chain-breaking process at the surface. The observations on the induction period, and the comparison between nitric oxide and propylene, confirm the alternative explanation in terms of the disturbing reaction.

When the induction period is short it is probable that the subsidiary surface reaction, which may be a condensation reaction of some sort, dies out very quickly. With *n*-heptane and *n*-pentane the steady rate seems to be independent of the induction period, and it may be assumed that the residual surface reaction is negligible.

With *n*-butane and propane, since the initial rate seems to vary somewhat with the induction period, it may be that under certain circumstances the rate of the disturbing reaction does not fall to zero, but rather to a small finite value which varies with the nature and extent of the surface. The present results for propane and *n*-butane are considered more nearly correct than those given in part II, since the induction

period was extremely short and the surface reaction probably negligible. Moreover, inhibition with propylene reduces the rates to the same limiting values.

The ratio of the rate in the standard vessel to that in the packed vessel being independent of the pressure, the variation of the activation energy with pressure would not itself be influenced by any surface effect even if some small amount were present. The difference between the higher and lower paraffins in this respect can not therefore be explained in any such way.

5 ANALYTICAL RESULTS

The breakage of the carbon-carbon linkages in the higher paraffin molecules can occur in several different places, and there is the possibility that the variation of activation energy with pressure might be connected with this circumstance. If reactions with different pressure dependence were involved in the breaking at the various places in the carbon chain, and if these were associated with distinct activation energies, then the overall value of E would become a function of the pressure.

As part of a more extended plan analysis of the products of the fully inhibited decomposition of *n*-heptane has been carried out at a series of temperatures, and it appears that the relative probability of breakage at the various points is more or less independent of temperature over the range in which E was measured. The results will be reported in detail elsewhere, but it may be concluded that the variation of activation energy with pressure cannot be ascribed to this cause.

6. INFLUENCE OF HYDROGEN

The question always present is whether the residual (maximally inhibited) reaction is a molecular rearrangement process or partly or wholly a chain reaction not suppressible by nitric oxide. One possibility which might be envisaged is a chain reaction involving hydrogen atoms, and immune to the action of nitric oxide in contrast to chains propagated by alkyl radicals which are suppressed. If there is a chain reaction of this type it will be common to all the paraffins, and the decomposition rate would, in general, be enhanced by hydrogen. Replacement of alkyl radicals by hydrogen atoms generated in the reaction $\text{CH}_3 + \text{H}_2 \rightarrow \text{CH}_4 + \text{H}^\cdot$ (activation energy 8.1 kcal.—Steacie 1947) would occur in competition with their removal by nitric oxide, thus more nitric oxide would be required to produce a given degree of inhibition.

The effect of increasing amounts of nitric oxide on the decomposition of 100 mm. propane at 570°C, alone, with 200 mm. and with 400 mm. of hydrogen, is shown in figure 5. Hydrogen causes a marked increase in the rate of the uninhibited decomposition, probably to a limiting value which is nearly double the original. The simplest interpretation of this effect is clearly that the chains are lengthened by the substitution of hydrogen atoms for methyl radicals according to the reaction mentioned above. These modified chains are still highly sensitive to the action of nitric oxide, 0.6 mm. of which reduces the rate to its normal value.

The limiting value to which the rate is lowered by sufficient nitric oxide is, however, very little changed by the hydrogen. What effect there is could easily be attributed to the normal contribution of the hydrogen molecules to the collisional activation process.

In any case the amount of nitric oxide required for a given proportional reduction of rate is nearly the same whether hydrogen is present or not. For the pure propane the amount required for a reduction to one-half is approximately 0.85 mm., and for the mixture with 400 mm. hydrogen it is approximately 0.7 mm.

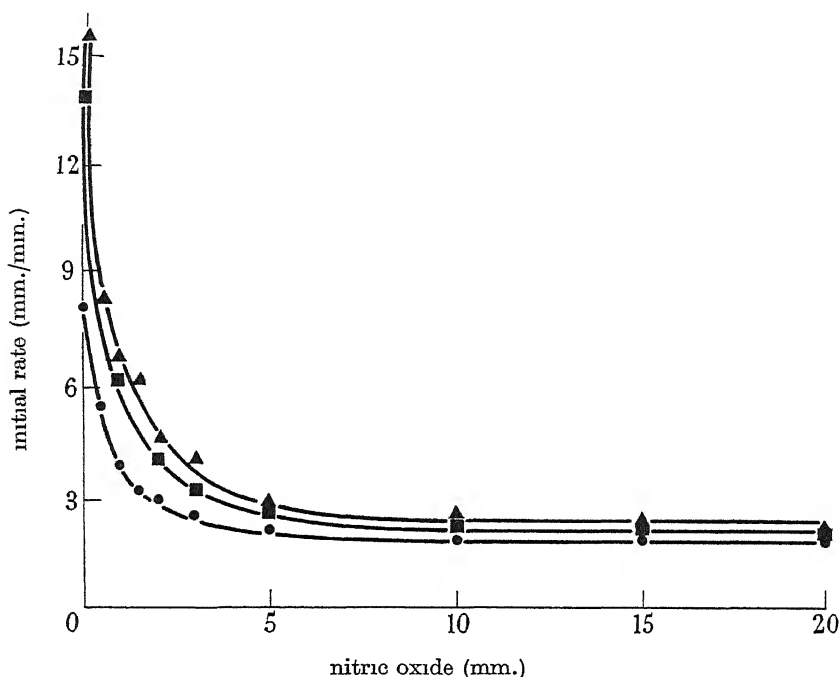


FIGURE 5. Thermal decomposition of 100 mm. propane at 570° C. ● without hydrogen; ■ 200 mm. hydrogen; ▲ 400 mm. hydrogen.

If the reaction between methyl radicals and hydrogen molecules has in fact an activation energy of 8.1 kcal., then at the reaction temperature one collision in 150 might give substitution of methyl by hydrogen. Since the pressure of hydrogen in the experiments is of the order 10^2 times that of the nitric oxide then, even on the most unfavourable assumption that methyl reacts with nitric oxide without activation energy, hydrogen and nitric oxide should compete for the methyl radicals on approximately equal terms. Thus if the hydrogen chains were in fact exempt from rupture by nitric oxide, inhibition should be more difficult in the presence of excess hydrogen.

Even qualitatively, the fact that the normal decomposition is much enhanced by hydrogen while this enhanced reaction shows very great sensitiveness to nitric oxide renders improbable the idea that the residual reaction involves hydrogen chains unsuppressible by this inhibitor.

7. ACTIVATION ENERGY AND PRESSURE

The present results indicate that the paraffins studied can be divided into two types. The activation energy for the higher paraffins (C_5 and C_7) is a function of the pressure, but this is not so for ethane and propane. Butane is intermediate in character.

By extrapolation of the plot of activation energy against pressure (figure 3) the value at zero pressure may be estimated. Similarly, a plot of activation energy against the reciprocal of the pressure gives the value at infinite pressure. For the higher paraffins, where the rate of the fully inhibited decomposition is given by equation (1) (§2), it can be seen that at low pressure the first-order term predominates, but at high pressure the second-order term predominates. Thus the limiting activation energies can also be obtained from the temperature coefficient of A and B . The values obtained in the two ways are recorded in table 5, and it is seen, particularly for n -pentane and n -heptane, that the agreement is fairly close. Steacie & Folkins (1940) found the activation energy at infinite pressure for the fully inhibited decomposition of n -butane to be 57.2 kcal., which is in agreement with the value obtained in the present investigation. With butane the extrapolated value of the activation energy at zero pressure does not agree very well with that obtained from the temperature coefficient of A , but, as mentioned previously, equation (1) (§2) does not apply very well to n -butane.

TABLE 5. LIMITING ACTIVATION ENERGIES (KCAL /G. MOL.)

	<i>n</i> -butane	<i>n</i> -pentane	<i>n</i> -heptane
by extrapolation to zero pressure	(70–80)	93	88
from temperature coefficient of A	69	93	87
by extrapolation to infinite pressure	58	63	62
from temperature coefficient of B	59	63	56

The form of the rate-pressure relation for propane is satisfied by equation (2) (§2), which is a combination of a first-order reaction (Ap_0), and a first-order reaction which in the pressure range investigated can become of the second order. The activation energy can be obtained by extrapolation to zero pressure, and the value of 67 kcal. obtained compares favourably with a value of 64 kcal. from the temperature coefficient of the constant A . Though extrapolation to infinite pressure will not give the value of the activation energy of the second reaction for propane (corresponding to $bp_0^2/(1+b'p_0)$ in equation (2)), this can be determined from the temperature coefficient of b . The value found is 7.45 kcal. (The activation energy corresponding to the temperature coefficient of b' was found to be about 13 kcal.)

Ethane presents a rather different case in that the activation energy does not vary with the pressure. The constant value of 74.7 kcal. is in excellent agreement with the value of 74.5 kcal. obtained at a single pressure by Staveley (1937), and in fairly good agreement with a value of 77.3 kcal. found by Steacie & Shane (1940).

Figure 2 of part II shows a plot of $\log r_\infty$ at 530°C against the number of carbon atoms, and it is seen that the relation is not quite a smooth one for the lower paraffins. The present investigation has shown that at a particular pressure there is no regular

variation of activation energy in the series ethane to heptane, and so there is no reason to expect a completely regular relation between $\log r_\infty$ and the number of carbon atoms. If the plot had been made for 600° C (with extrapolated values for *n*-heptane and *n*-pentane), the plot of $\log r_\infty$ against the number of carbon atoms would in fact have given a smooth curve.

8. GENERAL FORM OF THE RATE-PRESSURE RELATION

For the higher paraffins the results are given within the limits of experimental error by equation (1). With propane the second term reveals a tendency to change with increasing pressure from one representing a second-order reaction to one representing a first-order reaction. There being no reason whatever for supposing activation to be independent of collision, the first term should in all cases assume the second-order form at low enough pressures, though these are below the range which has been experimentally investigated. Moreover, the higher paraffins would at still greater pressures probably behave like propane, so that the general rate equation has in all probability the form

$$r_\infty = \frac{ap_0^2}{1+a'p_0} + \frac{bp_0^2}{1+b'p_0}, \quad (3)$$

which would correspond to a superposition of two unimolecular reactions with collisional activation. The constants of the first term are such that it assumes the first-order form at all pressures above a moderate value, those of the second such that it retains its second order form up to relatively high pressures.

With ethane the two terms appear to reduce to one.

9. GENERAL DISCUSSION

Any interpretation of these results must of course rest primarily upon a hypothesis about the nature of the residual reaction which is not suppressed by nitric oxide.

This, in principle, might be one of five things: (a) the primary process of what in the absence of inhibitors would be a chain reaction, (b) the normal chain reaction imperfectly suppressed by nitric oxide, (c) a surface reaction, (d) a second type of chain reaction not influenced by nitric oxide, (e) a molecular reaction.

Reasons for regarding (a) and (b) as improbable were given in part I, and for rejecting (c) in §4 of the present paper. No reasonable hypothesis of the type of (d) has yet appeared; the possibility that hydrogen atoms might propagate such a chain, distinct from the suppressible chains due to alkyl radicals, has not so far proved helpful (§6). While a hypothesis on these lines may be forthcoming in the light of further discoveries, in that of the present evidence (e) remains the most probable, and the results will therefore be considered on this basis.

For a single unimolecular reaction the rate is given by

$$r = \frac{k_1 p_0^2}{1 + (k_2/k_3) p_0}, \quad (4)$$

where k_1 is the rate constant of the activating process, k_2 that of the deactivating process, and k_3 , which need not be constant, depends upon the probability of the actual chemical transformation of the activated molecules.

k_3 can in general be a function of the total energy contained in the reacting molecule, in which case the pressure dependence of r assumes a more complicated form than that indicated by equation (4). The kind of dependence upon E normally envisaged for k_3 (Rice & Ramsperger 1927, Kassel 1928) could not, however, explain the dominance of the p_0^2 term at higher pressures, unless, as was pointed out in part II, the activated molecules fall into groups with very different transformation probabilities, corresponding to the two separate terms in equation (3).

As shown in table 5, the first term, which from its close approximation to a first power must correspond to a small transformation probability of activated molecules, is associated, as far as the paraffins above propane are concerned, with a very high activation energy. The second term, which corresponds to a much larger transformation probability, is associated with a much smaller activation energy.

What might be called a discrete spectrum of transformation probabilities is not unknown in decomposition reactions, and appears also from the results of Hunter (1934) on the thermal decomposition of nitrous oxide.

As stated in § 5, the different activation energies do not seem to be associated with different chemical modes of rupture of the molecule. They seem rather to depend upon physical differences in the activation mechanism.

TABLE 6. SQUARE-TERMS REQUIRED TO ACCOUNT FOR REACTION RATE

p_0 (mm.)	50	100	200	300	400	500
ethane (600°)	—	—	5	—	—	—
propane (580°)	9	8	8	7	7	6
<i>n</i> -butane (530°)	—	—	3	—	—	—
<i>n</i> -pentane (530°)	24	20	15	14	13	—
<i>n</i> -heptane (530°)	14	11	8	6	—	—

n is the minimum number of square terms required to account for the rate.

A hypothesis which appears worthy of serious consideration is the following. The reaction predominating at the lower pressures is one in which a large total quantity of energy E_1 enters the molecule and is spread over many degrees of freedom. If the molecule is left undisturbed, the interference of normal vibrations leads to a considerable accumulation of energy in one bond, or a very small number of bonds, and the molecule is disrupted. This so far is simply the standard theory of unimolecular reactions. Simultaneously, much more specifically defined collisions may place highly localized energy (corresponding to the lower activation energy E_2) into the molecule in such a way that very rapid transformation can occur. Collisions of this type correspond to the reaction which predominates at the higher pressures.

The first matter to consider in assessing such an hypothesis is whether the activation rates could be accounted for. Table 6 gives approximate values for the number of square terms in which, according to conventional methods of calculation, the activation energy would have to be distributed in order to account for the observed order of magnitude of the reaction rate. The numbers are well within what is permitted by the structural complexity of the molecules concerned. It is, perhaps, in favour of the view under discussion that the high activation energy mechanism appears only to be in play from butane upwards, where the necessary degrees of freedom are in fact present.

The next matter is whether the assumed activation modes are in fact physically plausible. As to the mechanism in which energy from very many square terms becomes concentrated in one or two bonds in such a way as to permit reaction, various authors have made calculations about the probability that with n quanta in k oscillators there should be j quanta in one of them, and if the molecule is large enough there is no particular difficulty about accounting for very considerable activation rates (Rice & Ramsperger 1927; Kassel 1928, Eley 1943; Evans & Rushbrook 1945; Barrer 1948).

In such calculations the quantity which appears in the reaction rate equation $k = A e^{-E/RT}$ is the total energy in all the degrees of freedom, that is to say a very large amount. On the other hand, because the molecule is so large, this quantity is not very many times the average energy for the temperature in question, so that the probability of its presence is quite considerable.

The detailed picture of its concentration in the bonds where it is chemically effective is usually left rather vague, and, at first sight, this process might seem to involve considerable elements of improbability. What must be remembered, however, is that the concentration need be a very fleeting one only, because it is followed by the irreversible chemical change. Suppose there were a molecule vibrating in five normal modes, and with a total energy of 100 kcal. shared between, say, five carbon-carbon bonds. This gives a rough average allotment of 4 kcal. per bond per mode. Sooner or later the amplitudes of the different modes are bound to reinforce one another at a given bond (though the maximum reinforcement is exceedingly transitory). This would give the certainty of E_0 in one bond for a time, which, although very brief, is long enough for chemical transformation to occur. A rearrangement reaction could easily be provoked by an amount of energy E_0 far smaller than that corresponding to the bond strength. The observed activation energy would, however, be not E_0 but the whole 100 kcal. This mechanism therefore is quite possible in principle.

We pass now to the mechanism in which much less energy than E_1 (though more than E_0) is placed in the molecule in a more localized manner. If the distribution is nearly correct at the moment of impact the transformation takes place almost immediately. Thus the probability of deactivation by collision is small. But the mode of activation here envisaged is itself useless unless the localization is nearly correct at the moment of impact, because the high concentration of energy in a few bonds which occurs at this moment will suffer an internal dissipation throughout the molecule unless the actual chemical transformation supervenes within a space of time short compared with the internal relaxation time. Once this internal dissipation has occurred the total energy E_2 is too small to give the re-accumulation by the interference of normal modes already considered, and which needs the greater quantity E_1 .

This view of the matter is perhaps the simplest which is in accord with the changes in the order of reaction over the range of pressure studied.

If a hypothesis on the above lines is entertained, the further question arises whether the sharp differentiation of the mechanisms is justified, or whether a whole range of behaviour between the two extremes should be postulated. The experimental results hardly allow a decision on this point. It raises, however, a problem for theoretical

solution; whether in fact the dynamics of molecules would be consistent with a separation of the two kinds of mechanism considered, or whether a range of intermediate cases would occur.

One of the authors (F.J.S.) is the holder of an Imperial Chemical Industries Research Fellowship. Another (K.U.I.) is in receipt of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

- Barrer, R. M. 1948 *Trans. Faraday Soc.* **44**, 399.
 Eley, D. D. 1943 *Trans. Faraday Soc.* **39**, 168.
 Evans, M. G. & Rushbrooke, G. S. 1945 *Trans. Faraday Soc.* **41**, 621.
 Hobbs, J. E. & Hinshelwood, C. N. 1938 *Proc. Roy. Soc. A*, **167**, 439.
 Hunter, E. 1934 *Proc. Roy. Soc. A*, **144**, 386.
 Kassel, L. S. 1928 *J. Phys. Chem.* **32**, 225.
 Partington, R. G., Stubbs, F. J. & Hinshelwood, Sir C. 1949 *J. Chem. Soc.* p. 2674.
 Rice, O. K. & Ramsperger, H. C. 1927 *J. Amer. Chem. Soc.* **49**, 1617.
 Staveley, L. A. K. 1937 *Proc. Roy. Soc. A*, **162**, 557.
 Steacie, E. W. R. 1947 *Disc. Faraday Soc.* 'The labile molecule', p. 80.
 Steacie, E. W. R. & Folkins, H. O. 1940 *Canad. J. Res. B*, **18**, 1.
 Steacie, E. W. R. & Shane, G. 1940 *Canad. J. Res. B*, **18**, 351.
 Stubbs, F. J. & Hinshelwood, Sir C. 1950a *Proc. Roy. Soc. A*, **200**, 458 (part I).
 Stubbs, F. J. & Hinshelwood, Sir C. 1950b *Proc. Roy. Soc. A*, **201**, 18 (part II).

Eigenfunction problems with periodic potentials

BY E. C. TITCHMARSH, F.R.S.

(Received 16 June 1950)

The expansion of an arbitrary function in terms of the solutions of the differential equation $\frac{d^2\phi}{dx^2} + \{\lambda - q(x)\} \phi = 0$ is obtained in the case where $q(x)$ is a periodic function of x .

1. The eigenfunction problem arising from the differential equation

$$\frac{d^2\phi}{dx^2} + \{\lambda - q(x)\} \phi = 0 \quad (-\infty < x < \infty), \quad (1.1)$$

where $q(x)$ is a periodic function of x , has been considered from various points of view by Hamel (1913), Haupt (1915, 1919), Strutt (1928), and Kronig & Penney (1930). The object of the present paper is to obtain the corresponding eigenfunction expansion of an arbitrary function $f(x)$ of integrable square.

The routine used in my book (Titchmarsh 1946) for solving such problems is as follows. Let $\Phi(x, \lambda)$ be the solution of the equation

$$\frac{d^2\Phi}{dx^2} + \{\lambda - q(x)\} \Phi = f(x). \quad (1.2)$$

To form this we construct solutions $\psi_1(x, \lambda)$ and $\psi_2(x, \lambda)$ of (1.1) whose moduli are of integrable square over $(-\infty, 0)$ and $(0, \infty)$ respectively. Then

$$\Phi(x, \lambda) = \frac{\psi_2(x, \lambda)}{\omega(\lambda)} \int_{-\infty}^x \psi_1(\xi, \lambda) f(\xi) d\xi + \frac{\psi_1(x, \lambda)}{\omega(\lambda)} \int_x^{\infty} \psi_2(\xi, \lambda) f(\xi) d\xi, \quad (1.3)$$

where $\omega(\lambda)$ is the Wronskian of ψ_1 and ψ_2 .

We then consider the integral

$$\frac{1}{2\pi i} \int \Phi(x, \lambda) d\lambda \quad (1.4)$$

taken round a large contour in the λ -plane. Under suitable conditions this tends to $f(x)$ as the contour tends to infinity. The expansion formula is then obtained by considering the singularities of $\Phi(x, \lambda)$.

2. We have first to obtain the functions ψ_1 and ψ_2 . Let a denote the period of $q(x)$, and suppose that $q(x)$ is continuous, or has a finite number of simple discontinuities in a period. Let $\theta(x, \lambda)$, $\phi(x, \lambda)$ be the solutions of (1.1) such that

$$\begin{aligned} \theta(0, \lambda) &= 1, & \theta'(0, \lambda) &= 0, \\ \phi(0, \lambda) &= 0, & \phi'(0, \lambda) &= -1, \end{aligned}$$

where dashes denote differentiations with respect to x . Then for any given x , $\theta(x, \lambda)$, $\theta'(x, \lambda)$, $\phi(x, \lambda)$ and $\phi'(x, \lambda)$ are integral functions of λ .

In the interval $na < x < (n+1)a$, where n is any integer, the general solution of (1.1) can be written as

$$\psi(x) = A_n \theta(x - na, \lambda) + B_n \phi(x - na, \lambda).$$

For continuity of $\psi(x)$ and $\psi'(x)$ at $x = na$, we must have

$$\begin{aligned} A_n &= A_{n-1} \theta(a, \lambda) + B_{n-1} \phi(a, \lambda), \\ B_n &= -A_{n-1} \theta'(a, \lambda) - B_{n-1} \phi'(a, \lambda). \end{aligned}$$

For brevity, let θ , θ' , ϕ , ϕ' denote $\theta(a, \lambda)$, $\theta'(a, \lambda)$, etc. Then

$$mA_n - B_n = (m\theta + \theta') A_{n-1} + (m\phi + \phi') B_{n-1},$$

where m is to be determined. If it is a root of the quadratic

$$m\theta + \theta' = -m(m\phi + \phi'), \quad (2.1)$$

then

$$mA_n - B_n = -(m\phi + \phi') (mA_{n-1} - B_{n-1}).$$

Hence, if m_1 and m_2 are the roots of (2.1),

$$\left. \begin{aligned} m_1 A_n - B_n &= (-m_1 \phi - \phi')^n (m_1 A_0 - B_0), \\ m_2 A_n - B_n &= (-m_2 \phi - \phi')^n (m_2 A_0 - B_0). \end{aligned} \right\} \quad (2.2)$$

If $m_1 \neq m_2$, these equations determine A_n and B_n in terms of A_0 and B_0 for all values of n .

If $B_0 = m_1 A_0$, then

$$B_n = m_1 A_n, \quad A_n = (-m_2 \phi - \phi')^n A_0 \quad (2.3)$$

for all values of n . If $B_0 = m_2 A_0$, then

$$B_n = m_2 A_n, \quad A_n = (-m_1 \phi - \phi')^n A_0 \quad (2.4)$$

for all values of n .

Now the equation (2.1) may be written

$$m^2 + \frac{\theta + \phi'}{\phi} m + \frac{\ddot{\phi}}{\phi} = 0,$$

and hence

$$\left(m + \frac{\theta + \phi'}{2\phi}\right)^2 = \frac{\theta^2 + \phi'^2 + 2\theta\phi' - 4\theta'\phi}{4\phi^2} = \frac{(\theta - \phi')^2 - 4}{4\phi^2},$$

since $\phi\theta' - \theta\phi' = \phi(0, \lambda)\theta'(0, \lambda) - \theta(0, \lambda)\phi'(0, \lambda) = 1$.

Hence
$$m = -\frac{\theta + \phi'}{2\phi} \pm \frac{\{(\theta - \phi')^2 - 4\}^{\frac{1}{2}}}{2\phi}. \quad (2.5)$$

Also
$$(m_1\phi + \phi')(m_2\phi + \phi') = \theta'\phi - \theta\phi' = 1. \quad (2.6)$$

Now as $|\lambda| \rightarrow \infty$ in the upper half-plane

$$\theta = \cos a\kappa + O\left(\frac{e^{a\tau}}{|\kappa|}\right), \quad \phi' = -\cos a\kappa + O\left(\frac{e^{a\tau}}{|\kappa|}\right) \quad (2.7)$$

where $\sqrt{\lambda} = \kappa = \sigma + i\tau$, $\sigma \geq 0$, $\tau \geq 0$ (see Titchmarsh 1946, (1.7.3), (1.7.7)). Hence

$$-m\phi - \phi' = \frac{1}{2}(\theta' - \phi) \pm \frac{1}{2}\{(\theta - \phi')^2 - 4\}^{\frac{1}{2}} = \cos a\kappa + O\left(\frac{e^{a\tau}}{|\kappa|}\right) \pm i\left\{\sin^2 a\kappa + O\left(\frac{e^{2a\tau}}{|\kappa|}\right)\right\}^{\frac{1}{2}}.$$

If $|\sin a\kappa| > A e^{a\tau}$ (i.e. we are not too near to any of the points $\kappa = n\pi/a$), this is equal to

$$\cos a\kappa \pm i \sin a\kappa + O(e^{a\tau}/|\kappa|).$$

For one root of the quadratic, say m_2 , we therefore obtain

$$-m_2\phi - \phi' = e^{-ia\kappa} + O(e^{a\tau}/|\kappa|). \quad (2.8)$$

It then follows from (2.6) that

$$-m_1\phi - \phi' = e^{ia\kappa} + O(e^{-a\tau}/|\kappa|). \quad (2.9)$$

The modulus of this expression is less than unity if $\tau > 0$ and $|\kappa|$ is large enough. Taking $A_0 = 1$, $B_0 = m_2$, we have

$$\psi_2(x, \lambda) = \theta(x, \lambda) + m_2(\lambda)\phi(x, \lambda),$$

and also

$$\psi_2(x, \lambda) = (-m_1\phi - \phi')^n \{\theta(x - na, \lambda) + m_2\phi(x - na, \lambda)\}, \quad (2.10)$$

and it is clear from this that ψ_2 is $L^2(0, \infty)$. Taking $A_0 = 1$, $B_0 = m_1$, we have

$$A_{-n} = (-m_2\phi - \phi')^{-n},$$

and the corresponding solution

$$\psi_1(x, \lambda) = \theta(x, \lambda) + m_1(\lambda)\phi(x, \lambda)$$

is $L^2(-\infty, 0)$.

3. The Wronskian of ψ_1 and ψ_2 is

$$\omega(\lambda) = m_1(\lambda) - m_2(\lambda) = \{(\theta - \phi')^2 - 4\}^{\frac{1}{2}}/\phi,$$

and the singularities of $\Phi(x, \lambda)$ are those of the numerator of this expression. Let

$$F_1(\lambda) = \theta(a, \lambda) - \phi'(a, \lambda) - 2 = \theta - \phi' - 2,$$

$$F_2(\lambda) = \theta(a, \lambda) - \phi'(a, \lambda) + 2 = \theta - \phi' + 2.$$

We shall see that in general each of these functions has an infinity of real zeros, which we shall denote by $\lambda_0, \lambda_1, \dots$ and $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots$ respectively. It will be shown that

$$\lambda_0 < \tilde{\lambda}_1 \leq \tilde{\lambda}_2 < \lambda_1 \leq \lambda_2 < \tilde{\lambda}_3 \leq \tilde{\lambda}_4 < \dots \quad (3.1)$$

Suppose, for example, that these numbers are all different. Then $\Phi(x, \lambda)$ has branch-points of the square-root type at each λ_n and $\tilde{\lambda}_n$, and the integral in (1.4) can be reduced to a series of loops, the first including λ_0 and $\tilde{\lambda}_1$, the second including $\tilde{\lambda}_2$ and λ_1 and so on.

Consider, for example, the contribution of the first loop to the expansion formula. The first term on the right-hand side of (1.3) is equal to

$$\begin{aligned} & \frac{\theta(x, \lambda) + m_2(\lambda) \phi(x, \lambda)}{m_1(\lambda) - m_2(\lambda)} \int_{-\infty}^x \{\theta(\xi, \lambda) + m_1(\lambda) \phi(\xi, \lambda)\} f(\xi) d\xi \\ &= \int_{-\infty}^x \left[\frac{\theta(x, \lambda) \theta(\xi, \lambda)}{m_1(\lambda) - m_2(\lambda)} + \frac{m_1(\lambda) m_2(\lambda) \phi(x, \lambda) \phi(\xi, \lambda)}{m_1(\lambda) - m_2(\lambda)} \right. \\ & \quad + \frac{1}{2} \frac{m_1(\lambda) + m_2(\lambda)}{m_1(\lambda) - m_2(\lambda)} \{\theta(x, \lambda) \phi(\xi, \lambda) + \phi(x, \lambda) \theta(\xi, \lambda)\} \\ & \quad \left. + \frac{1}{2} \theta(x, \lambda) \phi(\xi, \lambda) - \frac{1}{2} \phi(x, \lambda) \theta(\xi, \lambda) \right] f(\xi) d\xi. \end{aligned}$$

The last two terms are regular and so contribute nothing to the result. The others give

$$\begin{aligned} & \frac{1}{\{(\theta - \phi')^2 - 4\}^{\frac{1}{2}}} \int_{-\infty}^x [\phi \theta(x, \lambda) \theta(\xi, \lambda) + \theta' \phi(x, \lambda) \phi(\xi, \lambda) \\ & \quad - \frac{1}{2} (\theta + \phi') \{\theta(x, \lambda) \phi(\xi, \lambda) + \phi(x, \lambda) \theta(\xi, \lambda)\}] f(\xi) d\xi. \end{aligned} \quad (3.2)$$

Now $(\theta - \phi')^2 - 4$ is positive for large negative λ , and so in fact for $\lambda < \lambda_0$. As we pass above λ_0 , $\arg\{(\theta - \phi')^2 - 4\}$ decreases by π , and so $\{(\theta - \phi')^2 - 4\}^{\frac{1}{2}}$ becomes

$$-i\{4 - (\theta - \phi')^2\}^{\frac{1}{2}}.$$

We can reduce the loop to the straight line $(\lambda_0, \tilde{\lambda}_1)$ described twice, and the square root takes opposite values as we approach this segment of the real axis from above or below. Altogether these terms contribute to (1.4):

$$\begin{aligned} & -\frac{1}{\pi} \int_{\lambda_0}^{\tilde{\lambda}_1} \frac{d\lambda}{\{4 - (\theta - \phi')^2\}^{\frac{1}{2}}} \int_{-\infty}^x [\phi \theta(x, \lambda) \theta(\xi, \lambda) + \theta' \phi(x, \lambda) \phi(\xi, \lambda) \\ & \quad - \frac{1}{2} (\theta + \phi') \{\theta(x, \lambda) \phi(\xi, \lambda) + \phi(x, \lambda) \theta(\xi, \lambda)\}] f(\xi) d\xi. \end{aligned}$$

The second term on the right of (1.3) contributes a similar term with ξ varying over (x, ∞) . The other loops may be treated similarly. We thus obtain formally the expansion formula

$$\begin{aligned} f(x) &= -\frac{1}{\pi} \sum_{m=0}^{\infty} \int_{\lambda_{2m}}^{\tilde{\lambda}_{2m+1}} \frac{d\lambda}{\{4 - (\theta - \phi')^2\}^{\frac{1}{2}}} + \frac{1}{\pi} \sum_{m=1}^{\infty} \int_{\tilde{\lambda}_{2m}}^{\lambda_{2m-1}} \frac{d\lambda}{\{4 - (\theta - \phi')^2\}^{\frac{1}{2}}} \\ & \quad \times \int_{-\infty}^{\infty} [\phi \theta(x, \lambda) \theta(\xi, \lambda) + \theta' \phi(x, \lambda) \phi(\xi, \lambda) \\ & \quad - \frac{1}{2} (\theta + \phi') \{\theta(x, \lambda) \phi(\xi, \lambda) + \phi(x, \lambda) \theta(\xi, \lambda)\}] f(\xi) d\xi, \end{aligned} \quad (3.3)$$

where $\theta = \theta(a, \lambda)$, etc.

The spectrum, i.e. the set of values of λ which contribute to the expansion formula, thus consists of the intervals $(\lambda_{2m}, \tilde{\lambda}_{2m+1})$ and $(\tilde{\lambda}_{2m}, \lambda_{2m-1})$, which are in general separated by other intervals which do not belong to the spectrum. But if, for example, $\tilde{\lambda}_1 = \tilde{\lambda}_2$, then the first two intervals of the spectrum coalesce into one, this case is discussed in detail later.

To obtain the Parseval formula, multiply (3.3) by $f(x)$ and integrate with respect to x over $(-\infty, \infty)$. Writing

$$g(\lambda) = \int_{-\infty}^{\infty} \theta(x, \lambda) f(x) dx, \quad h(\lambda) = \int_{-\infty}^{\infty} \phi(x, \lambda) f(x) dx,$$

the result is

$$\int_{-\infty}^{\infty} \{f(x)\}^2 dx = -\frac{1}{\pi} \sum_{m=0}^{\infty} \int_{\lambda_{2m}}^{\tilde{\lambda}_{2m+1}} + \frac{1}{\pi} \sum_{m=1}^{\infty} \int_{\tilde{\lambda}_{2m}}^{\lambda_{2m-1}} \frac{\phi g^2(\lambda) + \theta' h^2(\lambda) - (\theta + \phi') g(\lambda) h(\lambda)}{\{4 - (\theta - \phi')^2\}^{\frac{1}{2}}} d\lambda. \quad (3.4)$$

This has the usual positive form since $\phi < 0$ in the intervals $(\lambda_{2m}, \tilde{\lambda}_{2m+1})$, and $\phi > 0$ in the intervals $(\tilde{\lambda}_{2m}, \lambda_{2m-1})$, while in either interval

$$4\phi\theta' - (\theta + \phi')^2 = 4(\phi\theta' - \phi'\theta) - (\theta - \phi')^2 = 4 - (\theta - \phi')^2 \geq 0.$$

4. Relations between the above problem and other eigenvalue problems

Consider the eigenvalue problem of the equation (1.1) taken over the finite interval $(0, a)$, with the boundary conditions

$$\psi(a) = \psi(0), \quad \psi'(a) = \psi'(0). \quad (4.1)$$

The general solution $A\theta(x, \lambda) + B\phi(x, \lambda)$ satisfies these boundary conditions if

$$A\theta + B\phi = A, \quad A\theta' + B\phi' = -B.$$

Hence

$$(\theta - 1)(\phi' + 1) = \theta'\phi,$$

$$\theta - \phi' = 1 + \theta'\phi - \theta\phi' = 2.$$

Hence the eigenvalues are the zeros $\lambda_0, \lambda_1, \dots$ of $F_1(\lambda)$. Also, at a zero λ_n of $F_1(\lambda)$, we have

$$m(\lambda) = -\frac{\theta + \phi'}{2\phi},$$

and so

$$-m\phi - \phi' = \frac{1}{2}(\theta - \phi') = 1.$$

Hence, in the notation of §2, $A_n = A_0$ and $B_n = B_0$, and so

$$\psi_2(x + a, \lambda_n) = \psi_2(x, \lambda_n), \quad \psi_1(x, \lambda) = \psi_2(x, \lambda).$$

The function $\psi_2(x, \lambda_n)$ is thus a multiple of the eigenfunction associated with λ_n for the boundary conditions (4.1). Similarly, the zeros $\tilde{\lambda}_n$ of $F_2(\lambda)$ are the eigenvalues of the corresponding problem with the boundary conditions

$$\psi(a) = -\psi(0), \quad \psi'(a) = -\psi'(0), \quad (4.2)$$

and

$$\psi_2(x + a, \tilde{\lambda}_n) = -\psi_2(x, \tilde{\lambda}_n).$$

The inequalities (3.1) now depend on 'oscillation theorems', i.e. theorems about the number of zeros in the interval $(0, a)$ of the eigenfunctions arising from the boundary conditions (4.1) or (4.2). Let $\psi(x)$, $\chi(x)$ be two such eigenfunctions, arising from either problem, $\psi(x)$ being associated with a smaller eigenvalue than $\chi(x)$. In the interval $(0, 2a)$, each has an even number of zeros, since $\psi(2a) = \psi(0)$, $\chi(2a) = \chi(0)$. Let $\psi(x)$ have $2k$ zeros in $(0, 2a)$. Then it follows from a theorem of Sturm (see Titchmarsh 1946, § 5.9) that $\chi(x)$ has at least $2k - 1$ zeros in $(0, 2a)$; and so, since the number is even, at least $2k$. Hence the number of zeros is a non-decreasing function of the eigenvalue.

According to a theorem of Birkhoff (1909) and Haupt (1915), the eigenfunction associated with λ_0 has no zeros, those associated with λ_{2m-1} and λ_{2m} ($m = 1, 2, \dots$) have $2m$ zeros in $(0, a)$, and those associated with $\tilde{\lambda}_{2m-1}$ and $\tilde{\lambda}_{2m}$ have $2m - 1$ zeros in $(0, a)$. From all these results, and the obvious fact that no λ_n is equal to any $\tilde{\lambda}_n$, the inequalities (3.1) follow.

Consider next the possibility that the sign of equality may occur in (3.1), i.e. that $F_1(\lambda)$ or $F_2(\lambda)$ may have a zero of higher order than the first. On differentiating (1.1) partially with respect to λ , we obtain

$$\frac{\partial^3 \phi(x, \lambda)}{\partial x^2 \partial \lambda} + \{\lambda - q(x)\} \frac{\partial \phi(x, \lambda)}{\partial \lambda} = -\phi(x, \lambda).$$

Also
$$\frac{\partial \phi(0, \lambda)}{\partial \lambda} = 0, \quad \frac{\partial \phi'(0, \lambda)}{\partial \lambda} = 0.$$

Hence
$$\frac{\partial \phi(x, \lambda)}{\partial \lambda} = -\int_0^x \{\theta(x, \lambda) \phi(\xi, \lambda) - \phi(x, \lambda) \theta(\xi, \lambda)\} \phi(\xi, \lambda) d\xi. \quad (4.3)$$

Similarly
$$\frac{\partial \theta(x, \lambda)}{\partial \lambda} = -\int_0^x \{\theta(x, \lambda) \phi(\xi, \lambda) - \phi(x, \lambda) \theta(\xi, \lambda)\} \theta(\xi, \lambda) d\xi. \quad (4.4)$$

Also, differentiating (4.3) with respect to x ,

$$\frac{\partial \phi'(x, \lambda)}{\partial \lambda} = -\int_0^x \{\theta'(x, \lambda) \phi(\xi, \lambda) - \phi'(x, \lambda) \theta(\xi, \lambda)\} \phi(\xi, \lambda) d\xi. \quad (4.5)$$

Putting $x = a$, and writing $\theta(a, \lambda) = \theta$, etc., we obtain

$$\frac{\partial}{\partial \lambda} \{\theta(a, \lambda) - \phi'(a, \lambda)\} = \int_0^a \{\phi \theta^2(\xi, \lambda) + \theta' \phi^2(\xi, \lambda) - (\theta + \phi') \theta(\xi, \lambda) \phi(\xi, \lambda)\} d\xi. \quad (4.6)$$

If $\phi \neq 0$, the right-hand side may be written

$$\int_0^a \left[\phi \left\{ \theta(\xi, \lambda) - \frac{\theta + \phi'}{2\phi} \phi(\xi, \lambda) \right\}^2 + \frac{4 - (\theta - \phi')^2}{4\phi} \phi^2(\xi, \lambda) \right] d\xi. \quad (4.7)$$

Suppose that λ is a zero of $F_1(\lambda)$ of the second order at least. Then (4.6) is zero, and if $\phi \neq 0$, (4.7) shows that

$$\theta(\xi, \lambda) = \frac{\theta + \phi'}{2\phi} \phi(\xi, \lambda) \quad (0 \leq \xi \leq a),$$

which is impossible. Hence $\phi = 0$; and similarly $\theta' = 0$. Since $\phi\theta' - \theta\phi' = 1$, it follows that $\theta\phi' = -1$. Hence

$$\theta - 2 = \phi' = -1/\theta, \quad (\theta - 1)^2 = 0,$$

and so $\theta = 1$, $\phi' = -1$.

The conditions

$$\theta(a, \lambda) = 1, \quad \theta'(a, \lambda) = 0, \quad \phi(a, \lambda) = 0, \quad \phi'(a, \lambda) = -1 \quad (4.8)$$

are therefore necessary for the occurrence of a zero of the second order (as least) of $F_1(\lambda)$; and from (4.6) they are plainly also sufficient.

Since
$$\frac{\partial^4 \phi(x, \lambda)}{\partial x^2 \partial \lambda^2} + \{\lambda - q(x)\} \frac{\partial^2 \phi(x, \lambda)}{\partial \lambda^2} = -2 \frac{\partial \phi(x, \lambda)}{\partial \lambda}$$

and
$$\frac{\partial^2 \phi(0, \lambda)}{\partial \lambda^2} = 0, \quad \frac{\partial^2 \phi'(0, \lambda)}{\partial \lambda^2} = 0,$$

we have
$$\frac{\partial^2 \phi(x, \lambda)}{\partial \lambda^2} = -2 \int_0^x \{\theta(x, \lambda) \phi(\xi, \lambda) - \phi(x, \lambda) \theta(\xi, \lambda)\} \frac{\partial \phi(\xi, \lambda)}{\partial \lambda} d\xi.$$

Similarly
$$\frac{\partial^2 \theta(x, \lambda)}{\partial \lambda^2} = -2 \int_0^x \{\theta(x, \lambda) \phi(\xi, \lambda) - \phi(x, \lambda) \theta(\xi, \lambda)\} \frac{\partial \theta(\xi, \lambda)}{\partial \lambda} d\xi.$$

If the conditions (4.8) hold, it follows that

$$\frac{\partial^2}{\partial \lambda^2} \{\theta(a, \lambda) - \phi'(a, \lambda)\} = -2 \int_0^a \left\{ \phi(\xi, \lambda) \frac{\partial \theta(\xi, \lambda)}{\partial \lambda} - \theta(\xi, \lambda) \frac{\partial \phi(\xi, \lambda)}{\partial \lambda} \right\} d\xi.$$

Replacing ξ by x , and then using (4.3) and (4.4), this becomes

$$-2 \int_0^a dx \int_0^x \{\theta(x, \lambda) \phi(\xi, \lambda) - \phi(x, \lambda) \theta(\xi, \lambda)\}^2 d\xi.$$

This clearly cannot vanish, so that $F_1(\lambda)$ cannot have a zero of higher order than the second.

Similarly, necessary and sufficient conditions that $F_2(\lambda)$ should have a zero of the second order are

$$\theta(a, \lambda) = -1, \quad \theta'(a, \lambda) = 0, \quad \phi(a, \lambda) = 0, \quad \phi'(a, \lambda) = 1. \quad (4.9)$$

It is now clear from (3.2) that $\Phi(x, \lambda)$ is regular at a double zero of $F_1(\lambda)$ or $F_2(\lambda)$.

5. Asymptotic formulae for $\psi_1(x, \lambda)$ and $\psi_2(x, \lambda)$

We have

$$\begin{aligned} \psi_2(a, \lambda) &= \theta + m_2 \phi = -m_1 \phi - \phi' \\ &= e^{i a \kappa} + O(e^{-a \tau} / |\kappa|) \end{aligned} \quad (5.1)$$

by (2.9), provided that $|\sin a \kappa| > A e^{a \tau}$. Since

$$\phi \psi_2'(a, \lambda) - \phi' \psi_2(a, \lambda) = \phi \theta' - \phi' \theta = 1,$$

and
$$\phi = -\frac{\sin a \kappa}{\kappa} + O\left(\frac{e^{a \tau}}{|\kappa|^2}\right), \quad (5.2)$$

we have, if $|\sin a \kappa| > A e^{a \tau}$,

$$\begin{aligned} \psi_2'(a, \lambda) &= \frac{1}{\phi} + \frac{\phi' \psi_2(a, \lambda)}{\phi} \\ &= \left(-\frac{\kappa}{\sin a \kappa} + \kappa \cot a \kappa e^{i a \kappa} \right) \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \\ &= i \kappa e^{i a \kappa} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\}. \end{aligned} \quad (5.3)$$

Next consider $\psi_2(x, \lambda)$ for $0 \leq x \leq a$. We have

$$\begin{aligned} \psi_2(x, \lambda) &= \psi_2(a, \lambda) \cos \kappa(x-a) + \psi_2'(a, \lambda) \frac{\sin \kappa(x-a)}{\kappa} \\ &\quad + \frac{1}{\kappa} \int_x^a \sin \kappa(\xi-x) q(\xi) \psi_2(\xi, \lambda) d\xi, \end{aligned} \quad (5.4)$$

e.g. by Titchmarsh (1946, lemma 1.7 (i)), with $\sin \alpha$ replaced by $\psi_2(a, \lambda)$ and $-\cos \alpha$ by $\psi_2'(a, \lambda)$. The first two terms on the right are

$$O(e^{-a\tau} e^{(a-x)\tau}) = O(e^{-x\tau}).$$

Hence
$$|\psi_2(x, \lambda)| \leq K e^{-x\tau} + \frac{K'}{|\kappa|} \int_x^a e^{(\xi-x)\tau} |\psi_2(\xi, \lambda)| d\xi,$$

where K, K' are independent of x and τ . Putting

$$|\psi_2(x, \lambda)| = e^{-x\tau} \Psi(x),$$

it follows that
$$\Psi(x) \leq K + \frac{K'}{|\kappa|} \int_x^a \Psi(\xi) d\xi,$$

and hence that $\Psi(x)$ is bounded for $|\kappa|$ large enough. Hence

$$\psi_2(x, \lambda) = O(e^{-x\tau}).$$

Hence (5.4) gives

$$\begin{aligned} \psi_2(x, \lambda) &= e^{ia\kappa} \cos \kappa(x-a) \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \\ &\quad + i\kappa e^{ia\kappa} \frac{\sin \kappa(x-a)}{\kappa} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} + O\left(\frac{1}{|\kappa|} \int_x^a e^{(\xi-x)\tau} e^{-\xi\tau} d\xi\right) \\ &= e^{ia\kappa} + O\left(\frac{e^{-x\tau}}{|\kappa|}\right) \quad (0 \leq x \leq a). \end{aligned} \quad (5.5)$$

Next consider values of x outside the interval $(0, a)$. Let

$$\chi(x) = (-m_1 \phi - \phi')^{-x/a} \psi_2(x, \lambda). \quad (5.6)$$

Then it follows from (2.10) that $\chi(x)$ has the period a . By (2.9) and (5.5)

$$\chi(x) = 1 + O\left(\frac{1}{|\kappa|}\right) \quad (5.7)$$

for $0 \leq x \leq a$, and so for all x . Hence

$$\begin{aligned} \psi_2(x, \lambda) &= \left[e^{ia\kappa} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \right]^{x/a} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \\ &= e^{ia\kappa} + O\left(\frac{e^{-x\tau}}{|\kappa|}\right) \end{aligned} \quad (5.8)$$

if x lies in any bounded interval.

The case of x unbounded requires further discussion. Consider the function

$$F(\lambda) = -m_1 \phi - \phi' = \frac{1}{2}(\theta - \phi') - \frac{1}{2}\{(\theta - \phi')^2 - 4\}^{\frac{1}{2}}$$

defined on the real axis by approach from above. For $\lambda < \lambda_0$, $\theta - \phi' > 2$, and the square root has its positive value. Hence this expression is of the form

$$c - \sqrt{(c^2 - 1)} = \frac{1}{c + \sqrt{(c^2 - 1)}} \quad (c > 1),$$

and so lies between 0 and 1.

For $\lambda_0 \leq \lambda \leq \tilde{\lambda}_1$,

$$-2 \leq \theta - \phi' \leq 2,$$

and

$$|F(\lambda)| = 1.$$

For $\tilde{\lambda}_1 < \lambda < \tilde{\lambda}_2$,

$$\theta - \phi' < -2,$$

and (since we have passed above branch-points at λ_0 and $\tilde{\lambda}_1$) the square root is real and negative. Hence $F(\lambda)$ is of the form

$$-c + \sqrt{c^2 - 1} \quad (c > 1),$$

and so lies between -1 and 0 .

Proceeding in this way, we see that

$$|F(\lambda)| \leq 1$$

on the whole real λ -axis. If we define $F(\lambda)$ in the lower half-plane by analytic continuation across the segment $(-\infty, \lambda_0)$ of the real axis, on which $F(\lambda)$ is real, the values which $F(\lambda)$ takes in the lower half-plane are the conjugates of those which it takes in the upper half-plane. In particular, $|F(\lambda)| \leq 1$ on the real axis approached from below.

Now suppose that $\lambda_0 \geq 0$, and consider the function $G(\lambda) = e^{-i\alpha\sqrt{\lambda}} F(\lambda)$. This is regular in the λ -plane cut along the real axis from 0 to infinity. On both sides of the cut, $|G(\lambda)| \leq 1$. Also by (2.9)

$$G(\lambda) = 1 + O(|\lambda|^{-\frac{1}{2}})$$

on certain large circles with centre at the origin. Hence, given $\epsilon > 0$,

$$|G(\lambda)| \leq 1 + \epsilon$$

on certain large circles. It follows from the maximum-modulus principle that the same inequality holds throughout the interior of any such circle; and so, since ϵ is arbitrary, $|G(\lambda)| \leq 1$, i.e. $|F(\lambda)| \leq e^{-\alpha\tau}$. Hence by (5.6) and (5.7)

$$|\psi_2(x, \lambda)| \leq e^{-x\tau} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \quad (5.9)$$

uniformly in x .

If $\lambda_0 < 0$, consider

$$G(\lambda) = e^{-i\alpha\sqrt{(\lambda-\lambda_0)}} F(\lambda).$$

This is regular in the plane cut from λ_0 to infinity. On the cut, $|G(\lambda)| \leq 1$; and on certain large circles

$$\begin{aligned} |G(\lambda)| &\leq |e^{i\alpha\{\sqrt{\lambda}-\sqrt{(\lambda-\lambda_0)}\}} \{1 + O(|\lambda|^{-\frac{1}{2}})\}| \\ &= 1 + O(|\lambda|^{-\frac{1}{2}}). \end{aligned}$$

It follows as before that $|G(\lambda)| \leq 1$. Hence

$$|F(\lambda)| \leq |e^{i\alpha\sqrt{(\lambda-\lambda_0)}}|$$

and

$$|F(\lambda)|^{x/a} \leq |e^{ix\sqrt{(\lambda-\lambda_0)}}| = e^{-x\tau} |e^{ix\{\sqrt{(\lambda-\lambda_0)}-\sqrt{\lambda}\}}|.$$

If $\tau \geq 1$, we deduce that

$$|F(\lambda)|^{x/a} \leq e^{-x\tau + Ax/|\kappa|},$$

and hence

$$|\psi_2(x, \lambda)| \leq e^{-x\tau + Ax/|\kappa|} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \quad (\tau \geq 1). \quad (5.10)$$

If $0 < \tau \leq 1$, $\sigma > 1$, we have

$$\begin{aligned}\sqrt{(\lambda - \lambda_0) - \sqrt{\lambda}} - \sqrt{\lambda} &= (\sigma^2 - \tau^2 + 2i\sigma\tau - \lambda_0)^{\frac{1}{2}} - \sigma - i\tau \\ &= (\sigma^2 - \lambda_0)^{\frac{1}{2}} \left(1 + \frac{2i\sigma\tau - \tau^2}{\sigma^2 - \lambda_0} \right)^{\frac{1}{2}} - \sigma - i\tau \\ &= (\sigma^2 - \lambda_0)^{\frac{1}{2}} + \frac{2i\sigma\tau - \tau^2}{2(\sigma^2 - \lambda_0)^{\frac{1}{2}}} + O\left(\frac{\tau^2}{\sigma}\right) - \sigma - i\tau.\end{aligned}$$

Hence

$$\begin{aligned}\Im\{\sqrt{(\lambda - \lambda_0) - \sqrt{\lambda}}\} &= \frac{\sigma\tau}{(\sigma^2 - \lambda_0)^{\frac{1}{2}}} - \tau + O\left(\frac{\tau^2}{\sigma}\right) \\ &= O\left(\frac{\tau}{\sigma^2}\right) + O\left(\frac{\tau^2}{\sigma}\right) = O\left(\frac{\tau}{\sigma}\right).\end{aligned}$$

Hence

$$|\psi_2(x, \lambda)| \leq e^{-x\tau + A x \tau / \sigma} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} \quad (0 < \tau \leq 1, \sigma > 1). \quad (5.11)$$

Also similar results hold for the function $\psi_1(x, \lambda)$, with x on the right-hand side replaced by $-x$. Finally, by (2.7) and (5.2),

$$\begin{aligned}m_1(\lambda) - m_2(\lambda) &= \{(\theta - \phi')^2 - 4\}^{\frac{1}{2}} / \phi \\ &= 2\kappa + O(1),\end{aligned} \quad (5.12)$$

provided that $|\sin a\kappa| > A e^{a\tau}$.

6. A convergence theorem

It will now be proved that the expansion (3.2) holds, in a certain sense, provided that the function $f(\xi)$ satisfies conditions similar to those under which Fourier's integral formula holds. *We shall assume that $|f(\xi)|$ is integrable over $(-\infty, \infty)$, and that $f(\xi)$ is of bounded variation in the neighbourhood of the point $\xi = x$.*

Consider the integral (1.4) taken round a contour in the λ -plane, symmetrical about the real axis, of which the upper half corresponds to the quarter-square in the κ -plane made up by the lines

$$\sigma = (n + \tfrac{1}{2})\frac{\pi}{a}, \quad 0 \leq \tau \leq (n + \tfrac{1}{2})\frac{\pi}{a}; \quad \tau = (n + \tfrac{1}{2})\frac{\pi}{a}, \quad 0 \leq \sigma \leq (n + \tfrac{1}{2})\frac{\pi}{a}.$$

Let

$$\begin{aligned}\Phi(x, \lambda) &= \frac{\psi_2(x, \lambda)}{m_1(\lambda) - m_2(\lambda)} \left(\int_{-\infty}^{x-\Delta} + \int_{x-\Delta}^{x-\delta} + \int_{x-\delta}^x \right) \psi_1(\xi, \lambda) f(\xi) d\xi \\ &\quad + \frac{\psi_1(x, \lambda)}{m_1(\lambda) - m_2(\lambda)} \left(\int_x^{x+\delta} + \int_{x+\delta}^{x+\Delta} + \int_{x+\Delta}^{\infty} \right) \psi_2(\xi, \lambda) f(\xi) d\xi \\ &= \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \Phi_5 + \Phi_6,\end{aligned}$$

where δ and Δ are to be determined. Let I_1, \dots, I_6 be the corresponding parts of the expression (1.4).

In the case $\lambda_0 \geq 0$ we obtain

$$\Phi_6 = O\left(\frac{e^{a\tau}}{|\kappa|} \int_{x+\Delta}^{\infty} e^{-\xi\tau} |f(\xi)| d\xi\right) = O\left(\frac{e^{-\Delta\tau}}{|\kappa|} \int_{x+\Delta}^{\infty} |f(\xi)| d\xi\right).$$

The last factor can be made arbitrarily small by choice of Δ , and, since

$$\int \Phi(x, \lambda) d\lambda = 2 \int \Phi(x, \kappa^2) \kappa d\kappa,$$

the first factor contributes to I_6

$$O\left(\int_0^{(n+\frac{1}{2})\pi/a} e^{-\Delta\tau} d\tau\right) + O\left(\int_0^{(n+\frac{1}{2})\pi/a} e^{-\Delta(n+\frac{1}{2})\pi/a} d\sigma\right) = O(1).$$

Hence I_6 can be made as small as we please, by choice of Δ , for all n .

In the case $\lambda_0 < 0$ we obtain

$$\begin{aligned} \Phi_6 &= O\left(\frac{e^{x\tau}}{|\kappa|} \int_{x+\Delta}^{\infty} e^{-\xi\tau + \Delta\xi/|\kappa|} |f(\xi)| d\xi\right) \\ &= O\left(\frac{1}{|\kappa|} e^{-\Delta\tau + \Delta(x+\delta)/|\kappa|} \int_{x+\Delta}^{\infty} |f(\xi)| d\xi\right) \quad (\tau \geq 1) \end{aligned}$$

and

$$\begin{aligned} \Phi_6 &= O\left(\frac{e^{x\tau}}{|\kappa|} \int_{x+\Delta}^{\infty} e^{-\xi\tau + \Delta\xi\tau/\sigma} |f(\xi)| d\xi\right) \\ &= O\left(\frac{1}{|\kappa|} e^{-\Delta\tau + \Delta(x+\Delta)\tau/\sigma} \int_{x+\Delta}^{\infty} |f(\xi)| d\xi\right) \quad (0 < \tau \leq 1). \end{aligned}$$

The same result clearly follows in this case.

Consider next Φ_4 , and choose δ so that $f(\xi)$ is of bounded variation in $x - \delta \leq \xi \leq x + \delta$.

Now

$$\begin{aligned} \Phi_4 &= \frac{e^{-ix\kappa}}{2\kappa} \int_x^{x+\delta} e^{i\xi\kappa} \left\{1 + O\left(\frac{1}{|\kappa|}\right)\right\} f(\xi) d\xi \\ &= \frac{e^{-ix\kappa}}{2\kappa} \int_x^{x+\delta} e^{i\xi\kappa} f(\xi) d\xi + O\left(\frac{1}{|\lambda|} \int_x^{x+\delta} |f(\xi)| d\xi\right). \end{aligned}$$

The last term contributes

$$O\left(\int_x^{x+\delta} |f(\xi)| d\xi\right)$$

to I_4 , and this can be made arbitrarily small by choice of δ . In the other term, if we replace $f(\xi)$ by $f(x+0)$, we obtain

$$f(x+0) \frac{e^{-ix\kappa}}{2\kappa} \int_x^{x+\delta} e^{i\xi\kappa} d\xi = f(x+0) \frac{e^{i\delta\kappa} - 1}{2i\kappa^2}.$$

It is easily seen that this contributes $\frac{1}{2}f(x+0) + o(1)$ to I_4 . Also

$$f(\xi) - f(x+0) = g_1(\xi) - g_2(\xi),$$

where g_1 and g_2 are positive non-decreasing functions of ξ , and we can suppose that $g_1(\xi) \leq \epsilon$, $g_2(\xi) \leq \epsilon$, where ϵ is given. Then by the second mean-value theorem

$$\begin{aligned} \int_x^{x+\delta} e^{-\xi\tau} \cos \xi\sigma g_1(\xi) d\xi &= g_1(x+\delta) \int_x^{x+\delta} e^{-\xi\tau} \cos \xi\sigma d\xi \\ &= g_1(x+\delta) \Re\left(\frac{e^{i\kappa(x+\delta)} - e^{i\kappa x}}{i\kappa}\right) = O\left(\frac{\epsilon e^{-x\tau}}{|\kappa|}\right). \end{aligned}$$

This contributes $O(\epsilon |\lambda|^{-1})$ to Φ_4 , and so $O(\epsilon)$ to I_4 ; and similarly for the other terms containing g_1 or g_2 .

Having fixed δ and Δ we can, in the interval $(x+\delta, x+\Delta)$, write

$$f(\xi) = g(\xi) + h(\xi),$$

where $g(\xi)$ is the integral of its derivative, and

$$\int_{x+\delta}^{x+\Delta} |h(\xi)| d\xi \leq \epsilon.$$

As in the case of I_6 , the contribution of $h(\xi)$ to I_5 is $O(\epsilon)$. That of $g(\xi)$ is

$$\begin{aligned} & \frac{e^{-i\lambda x}}{2\kappa} \int_{x+\delta}^{x+\Delta} e^{i\xi\kappa} \left\{ 1 + O\left(\frac{1}{|\kappa|}\right) \right\} g(\xi) d\xi \\ &= \frac{e^{-i\lambda x}}{2\kappa} \int_{x+\delta}^{x+\Delta} e^{i\xi\kappa} g(\xi) d\xi + O\left(\frac{e^{-\delta\tau}}{|\lambda|}\right) \\ &= \frac{e^{-i\lambda x}}{2\kappa} \left\{ \left[\frac{e^{i\xi\kappa}}{i\kappa} g(\xi) \right]_{x+\delta}^{x+\Delta} - \int_{x+\delta}^{x+\Delta} \frac{e^{i\xi\kappa}}{i\kappa} g'(\xi) d\xi \right\} + O\left(\frac{e^{-\delta\tau}}{|\lambda|}\right) \\ &= O\left(\frac{e^{-\delta\tau}}{|\lambda|}\right), \end{aligned}$$

which contributes $o(1)$ to I_5 .

Similar arguments apply to I_1 , I_2 and I_3 . Altogether we obtain

$$\lim_{\lambda \rightarrow \infty} \frac{1}{2\pi i} \int \Phi(x, \lambda) d\lambda = \frac{1}{2} \{f(x+0) + f(x-0)\}$$

under the above conditions. The integral on the left is, however, not equal to a finite sum of the terms on the right-hand side of (3.3), but to such a sum together with part of one term, e.g. an integral of the form

$$\int_{\lambda_{2m}}^{(n+\frac{1}{2})^2\pi^2/a^2} \dots d\lambda.$$

Whether this expression necessarily tends to zero is not clear.

Having proved the above convergence theorem, the Parseval theorem, that (3.4) holds for any function of integrable square, follows by well-known methods.

7. In this section we consider the relation between the above analysis and that of Kronig & Penney. These writers consider real values of λ only, and define the spectrum roughly as those λ -intervals in which periodic solutions of (1.1) are to be found. Now as in §5

$$\psi_2(x, \lambda) = (-m_1\phi - \phi')^{x/a} \chi(x),$$

where $\chi(x)$ has the period a . In the intervals of the spectrum,

$$|-m_1\phi - \phi'| = 1.$$

If

$$-m_1\phi - \phi' = e^{2\pi i h/k}, \tag{7.1}$$

where h and k are integers, then

$$(-m_1\phi - \phi')^{x/a} = e^{2\pi i h x/k a},$$

which has the period ka . Hence ψ_2 has the period ka . Values of λ for which (7.1) holds with some h and k are dense throughout the intervals of the spectrum, so that the spectrum as defined here coincides with that of Kronig & Penney.

Their first example is equivalent to taking

$$q(x) = 0 \quad (0 < x \leq b), \quad = c \quad (b < x \leq a).$$

This gives $\theta(x, \lambda) = \cos(x\sqrt{\lambda}), \quad \phi(x, \lambda) = -\frac{\sin(x\sqrt{\lambda})}{\sqrt{\lambda}}$

for $0 < x \leq b$, and

$$\theta(x, \lambda) = \cos(b\sqrt{\lambda}) \cos\{(x-b)\sqrt{(\lambda-c)}\} - \sqrt{\left(\frac{\lambda}{\lambda-c}\right)} \sin(b\sqrt{\lambda}) \sin\{(x-b)\sqrt{(\lambda-c)}\},$$

$$\phi(x, \lambda) = -\frac{\sin(b\sqrt{\lambda})}{\sqrt{\lambda}} \cos\{(x-b)\sqrt{(\lambda-c)}\} - \frac{\cos(b\sqrt{\lambda}) \sin\{(x-b)\sqrt{(\lambda-c)}\}}{\sqrt{(\lambda-c)}}$$

for $b < x \leq a$. The equations

$$\theta(a, \lambda) - \phi'(a, \lambda) = \pm 2$$

are therefore

$$2 \cos(b\sqrt{\lambda}) \cos\{(a-b)\sqrt{(\lambda-c)}\} - \left\{ \sqrt{\left(\frac{\lambda}{\lambda-c}\right)} + \sqrt{\left(\frac{\lambda-c}{\lambda}\right)} \right\} \sin(b\sqrt{\lambda}) \sin\{(a-b)\sqrt{(\lambda-c)}\} = \pm 2,$$

or

$$2 \cos\{b\sqrt{\lambda} + (a-b)\sqrt{(\lambda-c)}\} - \left\{ \sqrt{\left(\frac{\lambda}{\lambda-c}\right)} + \sqrt{\left(\frac{\lambda-c}{\lambda}\right)} - 2 \right\} \sin(b\sqrt{\lambda}) \sin\{(a-b)\sqrt{(\lambda-c)}\} = \pm 2.$$

Taking the + sign, the roots are given approximately by $a\sqrt{\lambda} \sim 2n\pi$. Putting

$$a\sqrt{\lambda} = 2n\pi + \delta,$$

it can be shown that

$$\delta = \frac{(a-b)ac}{4n\pi} \pm \frac{a^2c}{8n^2\pi^2} \sin \frac{2n\pi b}{a} + O\left(\frac{1}{n^3}\right).$$

A similar result is obtained on taking the negative sign. These formulae give the distribution of the intervals of the spectrum when λ is large.

The limiting case when $b \rightarrow a$, $c \rightarrow \infty$, in such a way that $(a-b)c \rightarrow k$, where k is a positive constant, is also considered. Here $\theta(x, \lambda)$ and $\phi(x, \lambda)$ are still continuous; but if $x_1 < b$, $x_2 > a$,

$$\begin{aligned} \phi'(x_2) - \phi'(x_1) &= \int_{x_1}^{x_2} \{q(x) - \lambda\} \phi(x) dx \\ &= k\phi(a) + O(x_2 - x_1) \end{aligned}$$

in the limiting case. Hence in this case

$$\phi'(a+0) - \phi'(a-0) = k\phi(a),$$

and similarly for $\theta(x, \lambda)$. The above analysis can be applied if we replace $\theta'(a, \lambda)$, $\phi'(a, \lambda)$ by $\theta'(a+0, \lambda)$, $\phi'(a+0, \lambda)$ respectively. Hence the branch-points are now given by the equations

$$2 \cos a\kappa + k \frac{\sin a\kappa}{\kappa} = \pm 2.$$

Taking the + sign, we obtain $a\kappa = 2n\pi$,

or
$$a\kappa = 2n\pi + \frac{ka}{2n\pi} + O\left(\frac{1}{n^2}\right),$$

and a similar result follows on taking the - sign.

REFERENCES

- Birkhoff, G. D. 1909 *Trans. Amer. Math. Soc.* **10**, 259.
 Hamel, G. 1913 *Math. Ann.* **73**, 371.
 Haupt, O. 1915 *Math. Ann.* **76**, 67.
 Haupt, O. 1919 *Math. Ann.* **79**, 278.
 Kronig, R. de L. & Penney, W. G. 1930 *Proc. Roy. Soc. A*, **130**, 499.
 Strutt, M. J. O. 1928 *Ann. Phys., Lpz.*, **86**, 319.
 Titchmarsh, E. C. 1946 *Eigenfunction expansions associated with second-order differential equations*. Oxford University Press

The cataphoresis of spherical, solid non-conducting particles in a symmetrical electrolyte

BY F. BOOTH

*H. H. Wills Physical Laboratory, Bristol University**

(Communicated by N. F. Mott, F.R.S.—Received 3 September 1948—

Revised 2 June 1950)

A theoretical analysis is given of the motion of solid non-conducting charged spheres through symmetrical electrolytes, under the action of an applied electric field. It is shown that the steady velocity of translation U may be written in the alternative forms

$$U = \sum_1^{\infty} c_\nu Q^\nu = \sum_1^{\infty} d_\nu \zeta^\nu,$$

where Qe denotes the charge on a sphere and ζ the zeta-potential; the coefficients c_ν and d_ν are proportional to the applied field strength and depend upon the radius of the particle and the concentrations, valencies and mobilities of the ions in the electrolyte. A general method is given for calculating the c_ν and d_ν and the first *four* coefficients of each series found explicitly. Some quite general properties of the c_ν are also deduced.

It is shown that under certain conditions, the terms of the series for U , apart from the first, are of considerable importance and must be taken into account in estimating ζ from U .

1. INTRODUCTION

In the following paper we shall consider the theory of one of the phenomena known as electrokinetic effects, namely, 'cataphoresis' or 'electrophoresis'. The effect we are to examine is the motion under the action of an electric field of charged solid spherical non-conducting particles suspended in water or in an electrolytic solution. Many attempts have been made in the past to determine the relation between the rate of motion of the particles, the applied field and other relevant physical quantities. Helmholtz (1879) was the first to make a theoretical study of electrokinetic phenomena in general, but he only gave a qualitative discussion of cataphoresis. Later Smoluchowski (1918) gave the formula

$$U = \frac{\epsilon E \zeta}{4\pi\eta} \tag{1.1}$$

* Present address: King's College, London.

for the cataphoretic velocity U , of non-conducting particles. ϵ denotes the dielectric constant of the liquid, E the applied field, η the coefficient of viscosity of the liquid. ζ represents the difference in potential between the surface of the solid and the liquid at infinity and is known as the 'electrokinetic potential' of the surface. It was assumed that both η and ϵ retained their macroscopic values in the immediate neighbourhood of the interface. The formula was claimed to be valid for particles of any size or shape provided that the thickness of the ionic atmosphere or 'double layer' at any point of the surface was small compared with the radius of curvature of the surface at that point.

The next major advance was made by Henry (1931). He confined attention to spherical particles but generalized Smoluchowski's work in two ways; he examined conducting particles and did not impose any restrictions on the double-layer thickness. For the case of non-conducting particles he gave the formula

$$U = \frac{\epsilon E \zeta}{4\pi\eta} f(b), \quad (1.2)$$

where $b = \chi a$, a is the particle radius and

$$\chi^2 = \frac{4\pi e^2 \sum_{i=1}^s n_i z_i^2}{\epsilon k T}. \quad (1.3)$$

$z_i e$ denotes the charge on an ion of type i , e being the electronic charge, n_i the concentration of ions of type i far from the particles; k is Boltzmann's constant and T the absolute temperature. The summation is over the s different ionic species in the electrolyte. The function $f(b)$ is plotted in figure 1 of Henry's paper; it increases monotonically from the value $\frac{2}{3}$ for small b , to 1 for large b . The quantity χ^{-1} has the dimensions of a length, and it is easily demonstrated that it is a rough measure of the thickness, normal to the surface, of the ionic atmosphere round the particle. Hence the parameter b represents the ratio of the linear dimensions of the particle to the double-layer thickness. The state $b \gg 1$ corresponds to a *thin* double layer, and we see that (1.2) then reduces to (1.1).

In deriving formula (1.2) Henry used two main simplifying assumptions. He assumed in the first place that the so-called 'inertia terms' in the hydrodynamical equations of motion could be neglected. Secondly, he regarded the field near the particle as simply the resultant of the applied field and the field due to the electrical double layer in its equilibrium state. The first assumption is almost certainly valid for the range of velocities usually encountered experimentally. But the second assumption is much more doubtful. As the particle moves through the electrolyte, the ions in the double layer will tend to lag behind. In front of the particle we would expect the charge density at a given position in the ionic atmosphere to be less than its equilibrium value, but behind the particle greater. This behaviour is well known in the Debye-Hückel theory of the conductance of electrolytes, where it is called the 'relaxation effect'. In addition, the applied field will modify the charge density in the atmosphere, quite apart from any effects due to the finite mobilities of the ions. It is the object of the present paper to remove this hiatus from Henry's work; we shall attempt to evaluate the potential near the particle exactly, and from this obtain

an improved formula for U . If we write the total charge on the particle as Qe , then we find that the cataphoretic velocity U may be expressed as a power series in Q ,

$$U = \sum_1^{\infty} c_{\nu} Q^{\nu}. \quad (1.4)$$

Since Q is also expressible as a function of the zeta potential, (1.4) can easily be rewritten as a power series in ζ . In what follows we shall develop a technique for the calculation of the coefficients c_{ν} . Unfortunately, the complexity of the evaluation increases rapidly with the order of ν , and consequently it has not been possible to work out the terms of series (1.4) explicitly, further than the fourth term. For the order of magnitude of the zeta potentials encountered experimentally, however, this should be adequate for a satisfactory estimate of ζ from measurements of U .

Three previous attempts have been made to improve Henry's work in the same respect, by Komagata (1935), by Hermans (1938) and by Overbeek (1943). The results of the first two papers are not in agreement and both treatments are invalid for different reasons. The major part of the present investigation was complete before the work of Overbeek came to the notice of the writer. Although the methods and scope of the present paper are different from Overbeek's, the results are in fair agreement.

2. BASIC ASSUMPTIONS AND METHOD OF CALCULATION

In what follows we shall consider non-conducting particles only. We make the following basic assumptions:

- (1) The dielectric constant ϵ , the viscosity η , and the ionic mobilities ω_i are uniform throughout the fluid phase.
- (2) The cataphoretic velocity is proportional to the field strength; this assumption simplifies the theory very much and is generally fulfilled in practice.
- (3) The electrolyte is incompressible.
- (4) The 'inertia terms' in the equations of motion of the electrolyte may be neglected. It can be shown that this is justifiable provided that (2) is fulfilled.
- (5) The electrolyte is *symmetrical*, that is, it contains equal numbers of ions of equal but opposite charge.
- (6) Finally, we require some assumptions on the behaviour at the surface of the particle. By purely thermodynamical arguments the difference of potential between the interior of the particle and the electrolyte far from the interface can be shown to be

$$\xi = \xi_0 + \frac{RT}{z_i F} \log_e a_i, \quad (2.1)$$

where ξ_0 is a constant, F is Faraday's constant, and a_i is the activity of the ion of type i . The potential ξ of (2.1) cannot be identified with the electrokinetic potential ζ , since ξ and ζ behave quite differently as the ionic concentrations are varied. It is necessary therefore to postulate a region or 'surface phase' on the solid side of the boundary, in which the potential varies. We shall make four assumptions on conditions in the surface phase:

- (i) The thickness of the surface phase is small compared with the particle radius.

(ii) The charge in the surface phase is immobile; there is no surface conductance effect.

(iii) Suppose that we denote the surface charge density when the field is applied, and when it is not applied, by $S(P)$ and $S_1(P)$ respectively. Now Henry assumed that

$$S(P) = -\frac{3E\epsilon_s}{8\pi} \cos \theta + S_1(P), \quad (2.2)$$

where ϵ_s is the dielectric constant of the solid, and θ the angle between the radius vector to P and the field direction. The first term represents a small charge due to the difference between the conductivities of solid and liquid. (Hence, apart from this small term, he assumed that the surface charge remained unchanged.) In our analysis we shall also assume (2.2); this does not imply of course that we shall go as far as Henry and postulate that the charge distribution in the *electrolyte* is also unchanged.

(iv) The potential difference across the surface phase retains its equilibrium value when the external field is applied.

Assumptions 6 (iii) and 6 (iv) are probably only approximations to the true state of affairs, since both the surface charge and potential jump must depend to some extent on conditions in the electrolyte—for example, on the ionic concentrations and the local field strength. It is hoped to examine this question in detail in a later paper. This completes the review of the basic assumptions, and we shall now proceed to sketch the method of calculation.

The motion of the fluid is steady if co-ordinates fixed with respect to the sphere are used. The equation of continuity becomes

$$\operatorname{div} \mathbf{v} = 0, \quad (2.3)$$

and, remembering assumptions (3) and (4) the equation of motion reduces to

$$\operatorname{grad} p + \rho \operatorname{grad} \psi = \eta \Delta \mathbf{v}. \quad (2.4)$$

\mathbf{v} is the fluid velocity, p the pressure, ρ the charge density and ψ the electrostatic potential.

If ω_i is the mobility of an ion (the velocity acquired when unit force is applied to it), then the mean velocity of the ions of type i is given by

$$\bar{\mathbf{u}}_i = \mathbf{v} - \omega_i \left(z_i e \operatorname{grad} \psi + \frac{kT}{m_i} \operatorname{grad} m_i \right), \quad (2.5)$$

where m_i is the concentration of the ions. The derivation of this equation is quite straightforward, the third term on the right-hand side represents the contribution to $\bar{\mathbf{u}}_i$ due to diffusion. The equation of continuity for the ions gives

$$\operatorname{div}(m_i \bar{\mathbf{u}}_i) = 0. \quad (2.6)$$

Equations (2.3), (2.4), (2.5) and (2.6), together with the boundary conditions, furnish all that is required to find the cataphoretic velocity, but a direct solution is impossible and we shall adopt the following indirect method.

(1) Equations (2.5) and (2.6) are solved with $\mathbf{v} = 0$ and $E = 0$. This solution for ψ , which we denote by ψ_1 , is the potential near the surface when the particle is at rest and no field is acting upon it.

(2) ψ_1 is substituted for ψ in (2.4), giving a solution \mathbf{v}_1 for the velocity. By adjusting the velocity at infinity so that the total force on the sphere vanishes, we obtain a first estimate U_1 for the cataphoretic velocity U .

(3) Suppose we now introduce \mathbf{v}_1 for \mathbf{v} in (2.5) and (2.6). The equations give a new solution for ψ which we can denote by ψ_2 .

(4) Repeating stage (2) with ψ set equal to ψ_2 instead of ψ_1 , we obtain a new value for U , namely, U_2 .

This alternate solution of the equations of motion of the electrolyte and of the ions can evidently be continued indefinitely, giving a series of successive approximations to the cataphoretic velocity. But if it proved necessary to carry through a large number of steps, the method would be very cumbersome, it is necessary therefore to consider the accuracy of the low-order approximations.

Let the total charge on the particle be Qe , where e denotes the electronic charge. Then we shall show that each approximation to U may be written in the form

$$U_\mu = \sum_{\nu=1}^{\infty} c_{\mu,\nu} Q^\nu. \quad (2.7)$$

The first suffix of the coefficient $c_{\mu,\nu}$ denotes the order of the approximation, the second the appropriate power of Q . Since Q can be expressed as a power series in ζ the electrokinetic potential, (2.7) can be written as

$$U_\mu = \sum_{\nu=1}^{\infty} d_{\mu,\nu} \zeta^\nu. \quad (2.8)$$

Now we shall find that

$$c_{1,1} = c_{2,1}, \quad c_{1,2} = c_{2,2}, \quad c_{1,\nu} \neq c_{2,\nu} \quad (\nu > 2). \quad (2.9)$$

Hence in the second-order approximation, the coefficients multiplying the first and second powers of Q are the same as in the first-order approximation; it is easily shown that the same applies for the higher order approximations, or symbolically,

$$c_{1,1} = c_{\mu,1}, \quad c_{1,2} = c_{\mu,2} \quad (\text{all } \mu). \quad (2.10)$$

Therefore, if the dependence of U on the first power of Q or ζ is required, it is sufficient to determine U_1 only; this is in effect Henry's method. Physically, this result indicates that the contribution to U due to the distortion of the ionic atmosphere, arising from the field and fluid motion, varies as powers of Q or ζ higher than the first. Hence it follows that Henry's third basic assumption, namely, 'that the applied field may be taken as simply superimposed on the field due to the electrical double layer' is valid for his result. The assumption would have been incorrect, however, if the dependence of U on higher powers of Q or ζ had been examined.

We shall also find that

$$c_{2,3} = c_{\mu,3}, \quad c_{2,4} = c_{\mu,4} \quad (\mu > 2). \quad (2.11)$$

Thus if we require the dependence of U on Q or ζ correct to the fourth power of either, we need only evaluate the second approximation. In general, U_n at least must be calculated to give the dependence of U on Q or ζ correct to the $2n$ th power. In the subsequent calculations we shall find U_1 and U_2 and therefore obtain the dependence of U on the first four powers of Q .

This completes our review of the method of calculation, and we can now proceed to determine the first approximation to ψ or ψ_1 .

3 CALCULATION OF THE FIRST APPROXIMATION TO THE POTENTIAL— NO APPLIED FIELD

Since equilibrium conditions obtain, we have

$$m_{1,i} = n_i \exp[-ez_i \psi_1 / kT]. \quad (3.1)$$

n_i is the concentration at a large distance from the particle of the ion of type i ; $m_{1,i}$ denotes the first approximation to the ionic concentration m_i . To determine ψ_1 we have, using Poisson's equation,

$$\Delta \psi_1 = -\frac{4\pi e}{\epsilon} \sum_{i=1}^s n_i z_i \exp[-ez_i \psi_1 / kT]. \quad (3.2)$$

The boundary conditions are

$$\psi_1 \rightarrow 0 \quad (r \rightarrow \infty), \quad \frac{d\psi_1}{dr} = -\frac{eQ}{\epsilon a^2} \quad (r \rightarrow a), \quad (3.3)$$

where r is the distance from the centre of the sphere. Equation (3.2) with a different second boundary condition has been solved by Gronwall, La Mer & Sandved (1928), and their method, with some slight modifications, can be used to find ψ_1 . If we set

$$\psi_1(x) = \sum_{\nu=1}^{\infty} e^{2\nu-1} (\epsilon a)^{-\nu} (-kT)^{1-\nu} Q^\nu \lambda_\nu(x), \quad (3.4)$$

where $x = \chi r$, then the functions $\lambda_\nu(x)$ are easily calculated:

$$\left. \begin{aligned} \lambda_1(x) &= \frac{b}{1+b} \frac{e^{b-x}}{x}, \quad \lambda_2(x) = 0, \\ \lambda_3(x) &= -\frac{q_3}{6} \left(\frac{b}{1+b} \right)^3 \frac{e^{3b}}{x} [\alpha(b) e^{-x} - 2e^x \text{Ei}(4x) + e^{-x} \text{Ei}(2x)], \\ \lambda_4(x) &= 0, \end{aligned} \right\} \quad (3.5)$$

$$\left. \begin{aligned} \text{where} \quad q_m &= \frac{\sum_{i=1}^s n_i z_i^{m+1}}{\sum_{i=1}^s n_i z_i^2}, \quad \text{Ei}(x) = \int_x^\infty \frac{e^{-u}}{u} du, \\ \alpha(b) &= \frac{e^{-2b}}{1+b} - \text{Ei}(2b) + 2 \left(\frac{1-b}{1+b} \right) e^{2b} \text{Ei}(4b). \end{aligned} \right\} \quad (3.6)$$

Equations (3.5) give all the λ_ν functions which will eventually be required.

4 CALCULATION OF THE FIRST-ORDER APPROXIMATION TO U

Henry calculated the cataphoretic velocity assuming that the potential due to the surface charge and the surrounding electrolyte retained spherical symmetry. Hence his results can be used immediately to determine U_1 , and we have from equation (18) of his paper

$$U_1 = \frac{\epsilon E}{6\pi\eta} \left[\psi_1(b) (1 + \sigma_0) + 3a^3 \sigma_0 \left(2 \int_a^\infty \frac{\psi_1(\chi r)}{r^4} dr - 5a^2 \int_a^\infty \frac{\psi_1(\chi r)}{r^6} dr \right) \right], \quad (4.1)$$

where $\sigma_0 = (\sigma' - \sigma)/(2\sigma' + \sigma)$, and σ' , σ are the specific conductivities of the electrolyte and solid respectively. Since in our case $\sigma = 0$, we have $\sigma_0 = \frac{1}{2}$. Changing to the x variable (4.1) becomes

$$U_1 = \frac{\epsilon E}{6\pi\eta} \left[\frac{3}{2} \psi_1(b) + 3b^3 \int_b^\infty \frac{\psi_1(x)}{x^4} dx - \frac{1}{2} b^5 \int_b^\infty \frac{\psi_1(x)}{x^6} dx \right]. \quad (4.2)$$

Substituting from series (3.4) for ψ_1 , and using the above expressions for the λ_ν , we obtain the series of coefficients $c_{1,1}, c_{1,2}, \dots$ for the expansion of U_1 in powers of Q .

$$c_{1,1} = \frac{Ee}{6\pi\eta a} X_1(b),$$

$$\text{with } X_1(b) = \frac{1}{(1+b)} \left[1 + \frac{1}{16}b^2 - \frac{5}{48}b^3 - \frac{1}{96}b^4 + \frac{1}{96}b^5 + \frac{1}{8}b^4 e^b \text{Ei}(b) \left(1 - \frac{1}{12}b^2 \right) \right],$$

$$c_{1,2} = 0,$$

$$c_{1,3} = \frac{Ee^5}{6\pi\eta a^3 (\epsilon kT)^2} q_3 X_3(b),$$

where

$$\begin{aligned} X_3(b) = & -\frac{1}{6} \left(\frac{b}{1+b} \right)^3 \left[\frac{\alpha(b) X_1(b) (1+b)}{b} + e^{4b} \text{Ei}(4b) \left(-\frac{2}{b} - \frac{1}{8}b - \frac{5}{24}b^2 + \frac{1}{48}b^3 + \frac{1}{48}b^4 \right) \right. \\ & - \frac{4}{3} b^3 e^{3b} \text{Ei}(3b) \left(1 - \frac{7}{30}b^2 \right) + e^{2b} \text{Ei}(2b) \left(\frac{1}{b} + \frac{1}{16}b - \frac{5}{48}b^2 - \frac{1}{96}b^3 + \frac{1}{96}b^4 \right) \\ & + \frac{1}{8}b^3 \left(1 - \frac{1}{12}b^2 \right) \left(\int_b^\infty \frac{e^{-x} \text{Ei}(2x)}{x} dx - 2 \int_b^\infty \frac{e^{-x} \text{Ei}(4x)}{x} dx \right) \\ & \left. - \frac{1}{48} \frac{1}{b} + \frac{3}{80} - \frac{9}{320}b + \frac{4}{960}b^2 + \frac{5}{1920}b^3 - \frac{7}{640}b^4 \right] \\ c_{1,4} = & 0. \end{aligned} \quad (4.3)$$

The function $X_3(b)$ is plotted in figure 1.

5. CALCULATION OF THE SECOND APPROXIMATION TO THE POTENTIAL

Following our general method we now proceed to solve equations (2.3), (2.5) and (2.6) for the second approximation ψ_2 to ψ , using in the equations the first approximation \mathbf{v}_1 , for the fluid velocity. From equations (2.5) and (2.6) we have

$$\text{div}(m_{2i}\mathbf{v}_1) = \omega_i \left[kT \Delta m_{2i} + ez_i \text{div} \left\{ m_{2i} \text{grad} \left(\psi_2 - rE \cos \theta \left[1 + \frac{a^3}{2r^3} \right] \right) \right\} \right], \quad (5.1)$$

where the m_{2i} denote the second approximations to the ionic concentrations m_i . Defining p_i and ϕ by the relations

$$m_{2i} = m_{1i} + p_i, \quad \psi_2 = \psi_1 + \phi, \quad (5.2)$$

$$\text{Poisson's equation gives} \quad \Delta \phi = -\frac{4\pi e}{\epsilon} \sum_{i=1}^s p_i z_i \quad (5.3)$$

and (5.1) becomes, after some rearrangement

$$\begin{aligned} kT \Delta p_i + ez_i \left[p_i (\Delta \psi_1 + \Delta \phi) + \left(\text{grad } p_i \text{grad} \left[\psi_1 + \phi - rE \cos \theta \left(1 + \frac{a^3}{2r^3} \right) \right] \right) \right] \\ + n_i ez_i \exp(-ez_i \psi_1 / kT) \left[\Delta \phi - \frac{ez_i}{kT} \frac{d\psi_1}{dr} \left(\frac{\partial \phi}{\partial r} - E \cos \theta \left[1 - \frac{a^3}{r^3} \right] \right) \right] \\ - \omega_i^{-1} \text{div} [\mathbf{v}_1 (n_i \exp(-ez_i \psi_1 / kT) + p_i)] = 0. \end{aligned} \quad (5.4)$$

To solve (5.3) and (5.4) we assume developments for p_i and ϕ in powers of Q

$$p_i = E \sum_{\nu=0}^{\infty} \sum_{n=0}^{\infty} e^{2\nu} (\epsilon kT)^{-\nu} a^{1-\nu} Q^{\nu} \pi_{i,\nu,n}(x) P_n(\cos \theta), \quad (5.5)$$

$$\phi = E \sum_{\nu=0}^{\infty} \sum_{n=0}^{\infty} e^{2\nu} (\epsilon kT)^{1-\nu} a^{1-\nu} Q^{\nu} \lambda_{\nu,n}(x) P_n(\cos \theta). \quad (5.6)$$

In accordance with our assumption that U varies linearly with the field strength, terms depending on E to powers other than the first are not included. From (5.3) the $\pi_{i,\nu,n}$ and $\lambda_{\nu,n}$ will be related

$$\Delta_n^{(r)} \lambda_{\nu,n} = -\frac{4\pi e}{\epsilon} \sum_{i=1}^s z_i \pi_{i,\nu,n}, \quad (5.7)$$

where $\Delta_n^{(r)}$ represents the operator

$$\Delta_n^{(r)} = \frac{d^2}{dr^2} + \frac{2}{r} \frac{d}{dr} - \frac{n(n+1)}{r^2}.$$

To determine p_i and ϕ we substitute expressions (5.5) and (5.6) in (5.4) and solve the differential equations for $\pi_{i,\nu,n}$ and $\lambda_{\nu,n}$ which are obtained by equating expressions multiplying different powers of Q to zero.

At infinity both $\pi_{i,\nu,n}$ and $\lambda_{\nu,n}$ are zero, the three boundary conditions at the surface are less simple. Since we have postulated no lateral movement of ions in the surface phase, it follows that the net flow of each type of ion towards the surface must vanish at all points on the surface. Hence if $\bar{u}_r(r, \theta)$ denotes the radial component of the average ionic velocity $\bar{\mathbf{u}}_i$ at the point r, θ , where θ is measured from the direction of E , one boundary condition is that

$$\bar{u}_r(a, \theta) = 0 \quad 0 \leq \theta \leq 2\pi.$$

From (2.5), since $v_r = 0$ at $r = a$, this becomes

$$kT \left. \frac{\partial m_i}{\partial r} \right|_a + e z_i m_i(a) \left. \frac{\partial \psi}{\partial r} \right|_a = 0. \quad (5.8)$$

The second boundary condition is derived from the assumptions (6) of § 2. Suppose we write the potential in the interior of the particle in the following way

$$\psi_{in} = -\frac{3}{2} \epsilon_s E r \cos \theta + \psi_s. \quad (5.9)$$

The first term is the potential the interior would have if the particle had no double layer. Just as we defined the sequence of approximations ψ_1, ψ_2, \dots for the potential in the double layer, we may write $\psi_{s,1}, \psi_{s,2}, \dots$ for the successive approximations to ψ_s corresponding to ψ_1 , etc. In this notation we can express 6 (iii) by the relation

$$\epsilon \left. \frac{\partial \phi}{\partial r} \right|_a - \epsilon_s \left. \frac{\partial \psi_{s,2}}{\partial r} \right|_a = 0 \quad (5.10)$$

and 6 (iv) by

$$\psi_{s,2}(a) - \xi - \phi(a) = 0. \quad (5.11)$$

We can now proceed to solve equation (5.4). Equating to zero the part independent of Q we find

$$\Delta_n^{(r)} (kT \pi_{i,0,n} + n_i e z_i \lambda_{0,n}) = 0. \quad (5.12)$$

Integrating we get

$$kT \pi_{i,0,n} + n_i e z_i \lambda_{0,n} = A_{i,0,n} r^n + B_{i,0,n} r^{-n-1}. \quad (5.13)$$

The boundary conditions give $A_{i,0,n} = B_{i,0,n} = 0$. (5.14)

Multiplying (5.13) by $4\pi e z_i / e k T$ and summing over the s ionic species, we find

$$(\Delta_n^{(r)} - \chi^2) \lambda_{0,n} = 0. \quad (5.15)$$

Hence $\lambda_{0,n}(x) = x^{-\frac{1}{2}} [C_{0,n} J_{n+\frac{1}{2}}(ix) + D_{0,n} J_{-n-\frac{1}{2}}(ix)]$, (5.16)

where $C_{0,n}$ and $D_{0,n}$ are integration constants and $J_n(x)$ is the usual Bessel function. From the condition that $\lambda_{0,n} \rightarrow 0$ as $r \rightarrow \infty$ the factor multiplying e^x in (5.16) must vanish; hence

$$D_{0,n} = -C_{0,n} (-i)^{-2n-1}. \quad (5.17)$$

From the solution for $\psi_{s,2}$ and the boundary conditions, we find

$$C_{0,n} \sum_{m=0}^n \frac{(n+m)!}{m!(n-m)!} \frac{1}{(2b)^m} [\epsilon(b+m+1) + n\epsilon_s] = 0. \quad (5.18)$$

Since each term of the summation is evidently positive, it follows that

$$C_{0,n} = 0. \quad (5.19)$$

Hence we conclude that for all n

$$\pi_{i,0,n} = \lambda_{0,n} = 0. \quad (5.20)$$

The first terms of series (5.5) and (5.6) for ϕ and p_i in powers of Q have now been found and we now proceed to examine the second terms. Equating the coefficient of Q to zero in (5.4)

$$\sum_{n=0}^{\infty} \Delta(kT\pi_{i,1,n} + n_i e z_i \lambda_{1,n}) P_n(\cos \theta) + \frac{n_i e z_i^2}{a} \left(1 - \frac{a^3}{r^3}\right) \cos \theta \frac{d\lambda_1}{dr} = 0. \quad (5.21)$$

This equation decomposes into two cases,

$$\Delta_n^{(r)}(kT\pi_{i,1,n} + n_i e z_i \lambda_{1,n}) = 0 \quad n \neq 1, \quad (5.22)$$

and

$$\Delta_1^{(r)}(kT\pi_{i,1,1} + n_i e z_i \lambda_{1,1}) = -\frac{n_i e z_i^2}{a} \left(1 - \frac{a^3}{r^3}\right) \frac{d\lambda_1}{dr}. \quad (5.23)$$

For $n \neq 1$, equation (5.22) is of exactly the same form as (5.12) for $\pi_{i,0,n}$ and $\lambda_{0,n}$ and may be analyzed in the same manner. We eventually find that, for $n \neq 1$,

$$\pi_{i,1,n} = \lambda_{1,n} = 0. \quad (5.24)$$

Now consider the solution of (5.23). Changing the independent variable from r to x and integrating

$$kT\pi_{i,1,1} + n_i e z_i \lambda_{1,1} = A_{i,1,1}x + \frac{B_{i,1,1}}{x^2} + n_i e z_i^2 F_1(b, x), \quad (5.25)$$

where $F_1(b, x) = \frac{e^{b-x}}{(1+b)} \left[\frac{1}{x} + \frac{1}{x^2} - \frac{b^3}{24} \left(-1 + \frac{1}{x} - \frac{2}{x^2} + \frac{6}{x^3} + x e^x \text{Ei}(x) \right) \right]$. (5.26)

The boundary condition at infinity gives $A_{i,1,1} = 0$. From the boundary condition (5.8) we obtain

$$B_{i,1,1} = \frac{1}{2} n_i e z_i^2 b^3 F_1'(b, b), \quad (5.27)$$

where

$$F_1'(b, b) = dF_1(b, x)/dx|_b. \quad (5.28)$$

For $\lambda_{1,1}$, since $q_2 = 0$,

$$(\Delta_1^{(x)} - 1) \lambda_{1,1}(x) = 0. \quad (5.29)$$

The appropriate solution is

$$\lambda_{1,1}(x) = C_{1,1} e^{-x} \left(\frac{1}{x} + \frac{1}{x^2} \right) + D_{1,1} e^x \left(\frac{1}{x} - \frac{1}{x^2} \right). \quad (5.30)$$

$$\text{From the boundary conditions} \quad C_{1,1} = D_{1,1} = 0. \quad (5.31)$$

Hence we finally conclude

$$\pi_{i,1,1} = \frac{n_i e z_i^2}{kT} \left[F_1(b, x) + \frac{b^3}{2x^2} F_1'(b, b) \right], \quad \lambda_{1,1} = 0. \quad (5.32)$$

The analysis for the determination of $\pi_{i,2,n}$, $\pi_{i,3,n}$, $\lambda_{2,n}$ and $\lambda_{3,n}$ is very similar to the cases just developed, we shall therefore omit details and merely state the results. We find

$$\pi_{i,2,n} = \lambda_{2,n} = 0 \quad n \neq 1, \quad (5.33)$$

and $\lambda_{2,1}$ satisfies the equation

$$\begin{aligned} (\Delta_1^{(x)} - 1) \lambda_{2,1}(x) = & -q_3 \left[F_2(b, x) + \frac{b^3}{2x^2} \left(F_2'(b, b) - \frac{1}{2} F_1'(b, b) - \frac{1}{b} F_1(b, b) \right) \right] \\ & - q_3^* \left[F_3(b, x) + \frac{b^3}{2x^2} F_3'(b, b) \right], \end{aligned} \quad (5.34)$$

where

$$q_m^* = \epsilon kT \sum_{i=1}^s n_i \omega_i^{-1} z_i^{m-1} \left/ e^2 \pi \eta \sum_{i=1}^s n_i z_i^2 \right. \quad (5.35)$$

The quantities $F_2(b, x)$ and $F_3(b, x)$ are, like $F_1(b, x)$, complicated functions of b and x . There is no point in carrying the integration of (5.34) a stage further since we shall eventually require the quantity $(\Delta_1^{(x)} - 1) \lambda_{2,1}$ only, not $\lambda_{2,1}$.

$$\text{Also} \quad \pi_{i,3,n} = \lambda_{3,n} = 0 \quad n \neq 1, \quad (5.36)$$

$$\text{and} \quad (\Delta_1^{(x)} - 1) \lambda_{3,1} = -q_4^* \left[F_6(b, x) + \frac{b^3}{2x^2} \left(F_6'(b, b) + \frac{d\lambda_1}{dx} \bigg|_b \left\{ F_3(b, b) + \frac{b}{2} F_3'(b, b) \right\} \right) \right]. \quad (5.37)$$

The necessary calculations of the second approximation to ψ are now complete and we can now find the second approximation to the fluid velocity.

6. CALCULATION OF THE SECOND APPROXIMATION FOR THE FLUID VELOCITY

We have for \mathbf{v}_2 and p_2 ,

$$\eta \operatorname{curl} \operatorname{curl} \mathbf{v}_2 + \operatorname{grad} p_2 = -\rho_2 \operatorname{grad} \left(\psi_2 - E \cos \theta \left[r + \frac{a^3}{2r^2} \right] \right). \quad (6.1)$$

$$\operatorname{div} \mathbf{v}_2 = 0. \quad (6.2)$$

$$\text{Assume that} \quad \psi_2 = \psi_1 + E \cos \theta \Phi, \quad (6.3)$$

where Φ is a function of r only. Also define p_2^* by

$$p_2^* = p_2 - \int_r^\infty \rho_1 \frac{d\psi_1}{dr} dr. \quad (6.4)$$

Then (6.1) becomes

$$\eta \operatorname{curl} \operatorname{curl} \mathbf{v}_2 + \operatorname{grad} p_2^* = \frac{\epsilon E}{4\pi} \left[\cos \theta \frac{d\psi_1}{dr} \Delta_1^{(r)} \Phi + \Delta_0^{(r)} \psi_1 \operatorname{grad} \left(\Phi - r - \frac{a^3}{2r^2} \right) \cos \theta \right] \quad (6.5)$$

if we drop terms involving E^2 in accordance with our basic assumptions.

The divergence of (6.5) gives an equation for Δp_2^* ; integrating,

$$p_2^* = \left(P_2 r + \frac{Q_2}{r^2} \right) \cos \theta + \frac{\epsilon E \cos \theta}{4\pi} \left[-\frac{d\psi_1}{dr} - a^3 r \int_r^\infty \frac{\Delta_0^{(z)} \psi_1}{z^4} dz \right. \\ \left. + \frac{1}{3} \int_r^\infty \left(2\Delta_0^{(z)} \psi_1 \Delta_1^{(z)} \Phi + \frac{d\psi_1}{dz} \frac{d(\Delta_1^{(z)} \Phi)}{dz} + \frac{d\Phi}{dz} \frac{d(\Delta_0^{(z)} \psi_1)}{dz} \right) \left(\frac{z^3}{r^2} - r \right) dz \right]. \quad (6.6)$$

From the boundary conditions at infinity $P_2 = 0$, but to find Q_2 the velocity must be investigated.

Let

$$\mathbf{w}_2 = \text{curl } \mathbf{v}_2. \quad (6.7)$$

In spherical polar co-ordinates (r, θ, α) , \mathbf{w}_2 has r and θ components equal to zero since there is symmetry about the axis $\theta = 0$, the same applies to $\text{curl curl } \mathbf{w}_2$. If we write for the α component of \mathbf{w}_2

$$\mathbf{w}_{2\alpha} = w^*(r) \sin \theta, \quad (6.8)$$

where $w^*(r)$ is a function of r only. Now it is easily proved from (6.8) that

$$\text{curl curl}_\alpha \mathbf{w}_2 = -\sin \theta \Delta_1^{(r)} w^*. \quad (6.9)$$

Using this relation we see that the curl of (6.5) gives an equation for $\Delta_1^{(r)} w^*$. The particular integral for w^* is

$$w^*(r) = -\frac{\epsilon E}{4\pi\eta} (\xi_1(r) + \xi_1^*(r)), \quad (6.10)$$

where

$$\xi_1(r) = \frac{d\psi_1}{dr} - \frac{a^3 r}{2} \int_r^\infty \frac{\Delta_0^{(z)} \psi_1}{z^4} dz, \quad (6.11)$$

$$\xi_1^*(r) = \frac{1}{3} \int_r^\infty \left(\frac{d\psi_1}{dz} \Delta_1^{(z)} \Phi - \Phi \frac{d(\Delta_0^{(z)} \psi_1)}{dz} \right) \left(\frac{z^2}{r^2} - \frac{r}{z} \right) dz. \quad (6.12)$$

Particular integrals for v_{2r} and $v_{2\theta}$ are readily found from equation (6.10). We omit details since the calculation is very nearly the same as that given by Henry, the only difference being the replacement of $\xi_1(z)$ by $\xi_1(z) + \xi_1^*(z)$. If $(-U_0)$ is the velocity at infinity, the complete solution is

$$\left. \begin{aligned} v_{2r} &= \cos \theta \left[U_0 \left(-1 + \frac{3a}{2r} - \frac{a^3}{2r^3} \right) + \frac{\epsilon E}{4\pi\eta} \left(\frac{a}{r} - \frac{a^3}{3r^3} \right) \int_a^\infty (\xi_1(z) + \xi_1^*(z)) dz \right. \\ &\quad \left. - \frac{\epsilon E}{6\pi\eta} \left(\int_r^\infty (\xi_1(z) + \xi_1^*(z)) dz + \frac{1}{r^3} \int_a^r (\xi_1(z) + \xi_1^*(z)) z^3 dz \right) \right], \\ v_{2\theta} &= \sin \theta \left[U_0 \left(1 - \frac{3a}{4r} - \frac{a^3}{4r^3} \right) - \frac{\epsilon E}{4\pi\eta} \left(\frac{a}{2r} + \frac{a^3}{6r^3} \right) \int_a^\infty (\xi_1(z) + \xi_1^*(z)) dz \right. \\ &\quad \left. + \frac{\epsilon E}{6\pi\eta} \left(\int_r^\infty (\xi_1(z) + \xi_1^*(z)) dz - \frac{1}{2r^3} \int_a^r (\xi_1(z) + \xi_1^*(z)) z^3 dz \right) \right], \\ v_{2\alpha} &= 0. \end{aligned} \right\} \quad (6.13)$$

$$p_2 = \int_r^\infty \rho_1 \frac{d\psi_1}{dz} dz + \cos \theta \left[\frac{3\eta a U_0}{2r^2} + \frac{\epsilon E}{4\pi} \left(\xi_1^{**}(r) + 2\xi_1(r) - 3 \frac{d\psi_1(r)}{dr} \right) \right. \\ \left. + \frac{\epsilon E}{4\pi r^2} \int_a^\infty (\xi_1(z) + \xi_1^*(z)) dz \right], \quad (6.14)$$

$$\text{where } \xi_1^{**}(r) = \frac{1}{3} \int_r^\infty \left[2\Delta_1^{(z)} \Phi \Delta_0^{(z)} \psi_1 + \frac{d\psi_1}{dz} \frac{d(\Delta_1^{(z)} \Phi)}{dz} + \frac{d\Phi}{dz} \frac{d(\Delta_0^{(z)} \psi_1)}{dz} \right] \left(\frac{z^3}{r^2} - r \right) dz. \quad (6.15)$$

To complete the solution it only remains to substitute

$$\Phi = \frac{e^4 Q^2}{a(\epsilon k T)^2} \lambda_{2,1} + \frac{e^6 Q^3}{a^2(\epsilon k T)^3} \lambda_{3,1} \quad (6.16)$$

in (6.12) and (6.15).

7. SECOND APPROXIMATION TO THE CATAPHORETIC VELOCITY

To determine the cataphoretic velocity we must now find the total force on the sphere, and choose the value of U_0 in equations (6.13) and (6.14) which makes this force vanish. This value of U_0 will then be the second approximation U_2 , to the cataphoretic velocity.

There are two kinds of forces to be considered:

(a) Electrostatic forces acting upon the charge on the surface. We shall denote the resultant electrostatic force by F_e .

(b) Surface stresses due to the liquid, giving a resultant force F_s on the sphere.

We shall first evaluate the electrostatic force on the sphere. The field in the electrolyte may be regarded as the sum of two fields, one with potential ψ_e , due to the external applied field together with the ions in the electrolyte, the other with a potential ψ_c due to the surface charge on the particle.

$$\psi = \psi_e + \psi_c. \quad (7.1)$$

To find the electrostatic force on the particle we require the force on the surface charge due to the part of the field derived from the potential ψ_e only.

The potential ψ_e is readily found. For, we have within the surface

$$\Delta \psi_e = 0. \quad (7.2)$$

The general solution of this equation is

$$\psi_e = \sum_{n=0}^{\infty} (M_n r^n + N_n r^{-n-1}) P_n(\cos \theta). \quad (7.3)$$

Since the region contains the origin we have

$$N_n = 0. \quad (7.4)$$

Also

$$M_n = 0 \quad n > 1, \quad (7.5)$$

since ψ involves $P_0(\cos \theta)$ and $P_1(\cos \theta)$ only, if we drop terms depending upon E^2 and higher powers. Hence the field in the direction of the θ axis is

$$X_e = - \frac{\partial \psi_e}{\partial (r \cos \theta)} \Big|_{\theta=0} = -M_1 \quad (7.6)$$

and is therefore uniform throughout the region concerned.

The field X_e is easily calculated since it is made up of the external field E and a term due to the asymmetry of the charge in the ionic atmosphere.

$$X_e = E - 2\pi \int_0^\pi \int_a^\infty \frac{\rho(r)}{r^2} \sin \theta \cos \theta r^2 d\theta dr. \quad (7.7)$$

Substituting from equation (6.3) the second integral becomes

$$\frac{1}{2}E \int_0^\pi \int_a^\infty \Delta_1^{(r)} \Phi(r) \sin \theta \cos^2 \theta d\theta dr. \quad (7.8)$$

Integrating, we eventually find that

$$X_e = E \left(1 - \frac{2}{3} \frac{\Phi(a)}{a} + \frac{1}{3} \frac{d\Phi(r)}{dr} \Big|_a \right). \quad (7.9)$$

Now the total electrostatic force on the sphere will simply be the total charge multiplied by the field X_e . Hence

$$F_e = eE \frac{d\psi_1}{dr} \Big|_a \left(\frac{a^2}{3} \frac{d\Phi(r)}{dr} \Big|_a + \frac{2a}{3} \Phi(a) - a^2 \right). \quad (7.10)$$

The force due to the pressure and viscous drag is easily found by integrating the stress components over the surface of sphere,

$$F_s = -6\pi a \eta U_0 - eEa \left[\frac{a}{3} \xi_1^{**}(a) - \frac{2a}{3} \xi_1^*(a) - a \frac{d\psi_1}{dr} \Big|_a + \int_a^\infty (\xi_1(z) + \xi_1^*(z)) dz \right]. \quad (7.11)$$

Now for U_0 and U_2 to be identical, the total force on the sphere must vanish. From this condition we obtain

$$U_2 = -\frac{eE}{6\pi\eta} \int_a^\infty (\xi_1(z) + \xi_1^*(z)) dz, \quad (7.12)$$

or, using (6.12) for $\xi_1^*(z)$

$$U_2 = U_1 - \frac{eE}{18\pi\eta} \int_a^\infty \int_r^\infty \left(\frac{d\psi_1(\chi z)}{dz} \Delta_1^{(z)} \Phi(z) - \Phi(z) \frac{d(\Delta_0^{(z)} \psi_1(\chi z))}{dz} \right) \left(\frac{z^2}{r^2} - \frac{r}{z} \right) dr dz. \quad (7.13)$$

Substituting for ψ_1 and Φ from (3.4) and (6.16) respectively we find, comparing coefficients of the first power of Q in equation (7.13),

$$c_{2,1} = c_{1,1}. \quad (7.14)$$

From (7.14) we observe that in the second approximation to U the first member of the series for U in powers of Q (which is the term calculated by Henry) is identical with the first member of the series in the first approximation.

Comparing coefficients of Q^2 we find

$$c_{2,2} = 0. \quad (7.15)$$

Since $c_{1,2} = 0$, we see that to order Q^2 , the first and second approximations agree. From the coefficients of Q^3 in (7.13) we find

$$c_{2,3} = c_{1,3} + \frac{Ee^5}{6\pi\eta a^3 (ekT)^2} [q_3 Y_3(b) + q_3^* Z_3(b)], \quad (7.16)$$

$$\text{where} \quad Y_3(b) = J(F_2, b) + \frac{b^3}{2} J\left(\frac{1}{z^2}, b\right) [F'_2(b, b) - \frac{1}{2}F'_1(b, b) - \frac{1}{2}F_1(b, b)], \quad (7.17)$$

$$\text{and} \quad Z_3(b) = J(F_3, b) + \frac{b^3}{2} J\left(\frac{1}{z^2}, b\right) F'_3(b, b). \quad (7.18)$$

In (7.17) and (7.18) the symbol $J(f, b)$ denotes the integral operator defined by

$$J(f(z), b) = \frac{b^2 e^b}{3(1+b)} \int_b^\infty \int_x^\infty e^{-z} f(z) \left(1 + \frac{1}{z}\right) \left(\frac{x}{z^2} - \frac{z}{x^2}\right) dx dz. \quad (7.19)$$

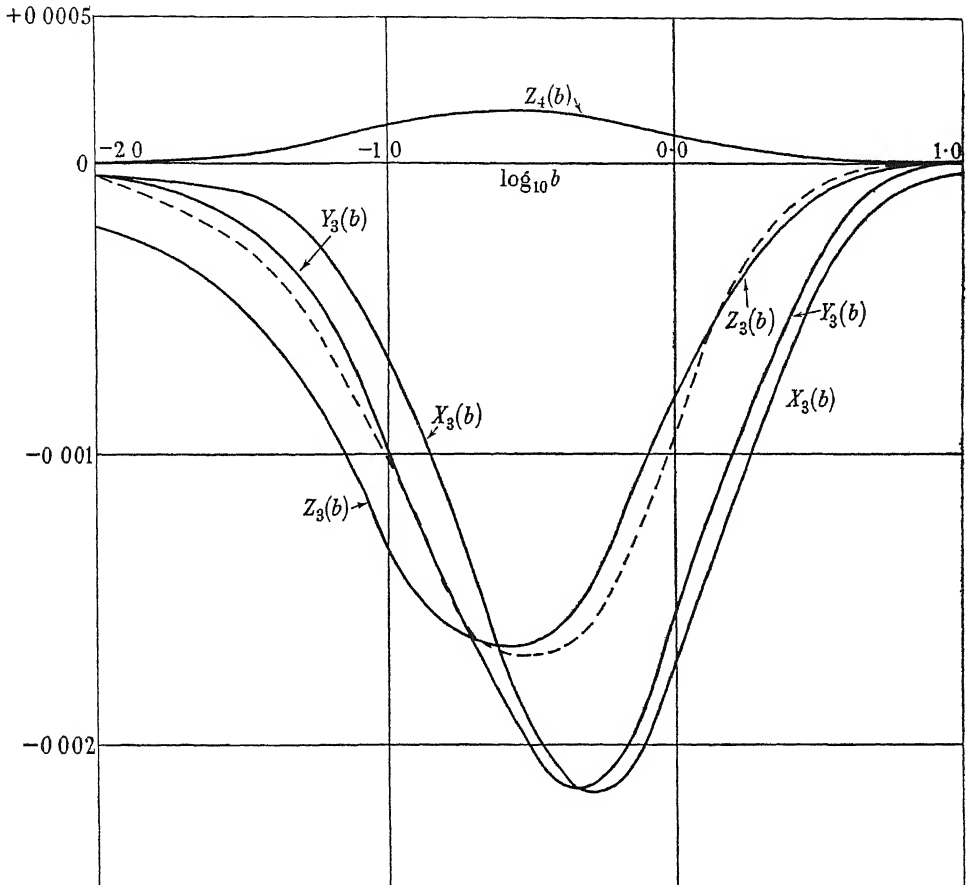


FIGURE 1. Functions $X_3(b)$, $Y_3(b)$, $Z_3(b)$ and $Z_4(b)$. The dashed curve is $-10Z_4(b)$.†

$Z_3(b)$ (apart from a multiplying factor) represents the effect of the distortion of the double-layer due to the 'relaxation effect' on $c_{2,3}$, whereas $Y_3(b)$ represents the effect of the remaining part of the distortion. The evaluation of (7.17) and (7.18) is very tedious; $Y_3(b)$ and $Z_3(b)$ are plotted however in figure 1. Both functions are everywhere negative, and tend to zero for both small and large b .

† The author will supply the complete analytical expressions for $X_3(b)$, $Y_3(b)$, $Z_3(b)$ and $Z_4(b)$ to any readers who might desire them.

Returning to equation (7.13) and comparing coefficients of Q^4 we find

$$c_{2,4} = \frac{Ee^7}{6\pi\eta a^4(ekT)^3} q_4^* Z_4(b), \quad (7.20)$$

where
$$Z_4(b) = J(F_6, b) + \frac{b^3}{2} J\left(\frac{1}{z^2}, b\right) \left[F_6'(b, b) - \frac{1}{2} F_3'(b, b) - \frac{1}{b} F_3(b, b) \right]. \quad (7.21)$$

We shall omit the analytical expression for $Z_4(b)$, since it is extremely complicated; the function is plotted in figure 1. It is everywhere positive, with a single maximum and vanishes for both small values of b and large ones. For convenience in calculation we have also plotted the quantity $-10Z_4(b)$, since the $Z_4(b)$ curve lies near the axis of $\log_{10} b$ on the scale of figure 1.

This completes the information we can derive from the results of §5, and we must now proceed to higher approximations.

8. HIGHER APPROXIMATIONS

Following our general plan the next step in our calculation would be to find the third approximation ψ_3 to ψ from equation (5.1) with \mathbf{v}_1 replaced by \mathbf{v}_2 , the second approximation to the fluid velocity derived from equation (6.13). From ψ_3 we could then find \mathbf{v}_3 and so U_3 by the method of the preceding paragraph. Unfortunately, owing to the elaborate expression for \mathbf{v}_2 the method becomes prohibitively complicated for approximations higher than the second. However, it is possible by very general arguments to determine some useful properties of the higher coefficients of series (1.4).

The second approximation U_2 gives U correct to the fourth power of Q or ζ . For from (6.13), (6.12) and (6.16) \mathbf{v}_2 differs from \mathbf{v}_1 in terms of order Q^3 . Hence from (5.1) the difference between ψ_3 and ψ_2 is of order Q^4 , and from (6.12) and (7.13) U_3 and U_2 differ by terms of order Q^5 . Similarly, we can show that U_4 and U_3 differ by terms of order Q^7 at most. Continuing the argument the statements of the second paragraph on the $c_{n,\nu}$ coefficients can easily be verified.

Although we shall not attempt to find the coefficients c_n for $n > 4$ explicitly, we shall now prove that their limiting values for b very large or very small are all zero,

$$c_\nu = 0 \quad b \ll 1 \quad \text{or} \quad b \gg 1 \quad (\nu \neq 1). \quad (8.1)$$

First consider the case of $b \ll 1$; since χ^{-1} is a measure of the double layer thickness the case $b \ll 1$ corresponds to an ionic atmosphere of infinite thickness. Now the total charge in the ionic atmosphere must be equal and opposite to the charge on the surface. Hence a double layer of infinite thickness will have a charge so diffuse that its effect on the particle may be ignored. The system is therefore equivalent to a charged particle in a non-ionizing medium. In this case the velocity is obviously proportional to the charge Q , since the viscous drag is simply given by the Stokes formula. This proves the first part of assertion (8.1).

For the other case, $b \gg 1$, it is convenient to consider separately the contributions to the coefficients c_ν of the spherically symmetrical and spherically unsymmetrical parts of the electrostatic field.

Since it can easily be shown by the methods of paragraph 4 that all the $\lambda_{\nu,n}$ are zero except when $n = 1$, it follows that the contribution to c_ν due to the symmetrical part of the field is simply $c_{1,\nu}$. From equations (3.4) and (4.2) we have

$$c_{1,\nu} = \frac{Ee^{2\nu-1}}{6\pi\eta a^\nu (-\epsilon kT)^{\nu-1}} \left[\frac{3}{2} \lambda_\nu(b) + 3b^3 \int_b^\infty \frac{\lambda_\nu(x)}{x^4} dx - \frac{15}{2} b^5 \int_b^\infty \frac{\lambda_\nu(x)}{x^6} dx \right]. \quad (8.2)$$

We therefore require the behaviour of $\lambda_\nu(x)$ for large values of both x and b . Let us assume that for $b \gg 1$, $\lambda_\nu(x)$ can be written in the form

$$\lambda_\nu(x) = \sum_{n=1}^{\nu} a_{\nu,n} \frac{e^{n(b-x)}}{b^{\nu-n} x^n} \left[1 + O\left(\frac{1}{b}\right) \right], \quad (8.3)$$

where the coefficients $a_{\nu,n}$ are independent of x and b , being products of ordinary numbers and the quantities q_1, q_2, \dots . It is easily verified that the $\lambda_\nu(x)$ functions which we have worked out explicitly are of this form.

Now from $\lambda_\nu(x)$ the higher functions $\lambda_{\nu+1}(x)$, $\lambda_{\nu+2}(x)$ can be found by a fairly straightforward application of the Gronwall, La Mer, Sandved method. Using this, it is not difficult to show that $\lambda_{\nu+1}(x)$ can be put into the form of (8.3). Hence by induction it follows that (8.3) is valid. Substituting for $\lambda_\nu(x)$ in equation (8.2)

$$c_{1,\nu} = \frac{Ee^{2\nu-1}}{6\pi\eta a^\nu (-\epsilon kT)^{\nu-1}} \left[\frac{3}{2b^\nu} + O\left(\frac{1}{b^{\nu+1}}\right) \right] \sum_{n=1}^{\nu+1} a_{\nu,n}. \quad (8.4)$$

Hence the 'symmetrical field' part of the coefficient of Q in the series development for U tends to zero as $b^{-\nu}$ for large b .

The contribution to U from the asymmetrical part of the double layer field can be investigated in a rather similar fashion, using the analysis of §4. We shall omit details of the rather tedious argument and merely state the conclusion. We find

$$c_\nu - c_{1,\nu} = O(b^{-2}) \quad (8.5)$$

at most. We finally reach the conclusion therefore that the second part of assertion (8.1) is correct.

9. SUMMARY AND DISCUSSION

It is now convenient to collect our results and discuss them. We may evidently write the general formula for U in the form

$$U = \frac{EeQ}{6\pi\eta a} X_1(b) + \frac{E}{6\pi\eta} \sum_{\nu=2}^{\infty} Q^\nu e^{2\nu-1} a^{-\nu} (\epsilon kT)^{1-\nu} \\ \times [X_\nu^\dagger(b; q_2, \dots, q_\nu) + Y_\nu^\dagger(b; q_2, q_3, \dots, q_\nu) + Z_\nu^\dagger(b; q_2, q_3, \dots, q_\nu, q_3^*, \dots, q_\nu^*)]. \quad (9.1)$$

The first term is simply Henry's formula. The physical significance of the three functions X_ν^\dagger , Y_ν^\dagger , and Z_ν^\dagger in the second term is as follows: X_ν^\dagger represents a contribution to U due to the *symmetrical* part of the field round the particle. These terms did not appear in Henry's formula because the Debye-Hückel approximation was invoked. Both Y_ν^\dagger and Z_ν^\dagger represent contributions due to distortion of the

double-layer field, but Y_v^\dagger is independent of the ionic mobilities. For symmetrical electrolytes we have found

$$X_2^\dagger = Y_2^\dagger = Z_2^\dagger = X_4^\dagger = Y_4^\dagger = 0, \quad (9.2)$$

$$\left. \begin{aligned} X_3^\dagger(b, q_2, q_3) &= q_3 X_3(b) \\ Y_3^\dagger(b; q_2, q_3) &= q_3 Y_3(b) \\ Z_3^\dagger(b; q_2, q_3, q_3^*) &= q_3^* Z_3(b) \\ Z_4^\dagger(b; q_2, q_3, q_4; q_3^*, q_4^*) &= q_4^* Z_4(b). \end{aligned} \right\} \quad (9.3)$$

where $X_3(b)$, $Y_3(b)$, $Z_3(b)$ and $Z_4(b)$ are functions of the parameter 'b' only, and are plotted in figure 1.

An alternative expression for U is an expansion in powers of ζ .

$$U = \frac{Ee\zeta}{6\pi\eta} X_1^*(b) + \frac{EekT}{6\pi\eta e} \sum_{v=2}^{\infty} \left(\frac{e\zeta}{kT} \right)^v \times [X_v^s(b; q_2, \dots, q_v) + Y_v^s(b; q_2, \dots, q_v) + Z_v^s(b; q_2, \dots, q_v; q_3^*, \dots, q_v^*)]. \quad (9.4)$$

The physical significance of the various terms X_v^s , Y_v^s , Z_v^s corresponds to that of the X^\dagger , Y^\dagger and Z^\dagger of (9.1). A set of relations exactly analogous to (9.2) and (9.3) holds for the X_v^s , Y_v^s , Z_v^s functions, the unstarred functions of b on the right-hand side of (9.3) being replaced by starred functions $X_3^*(b)$, $Y_3^*(b)$, $Z_3^*(b)$ and $Z_4^*(b)$; these quantities are plotted in figure 2. $X_1^*(b)$ is plotted by Henry (figure 1). It will be observed from figure 2 that the starred functions plotted tend to zero as b becomes infinite. Unfortunately it cannot be inferred from §8 that *all* the functions X_v^s , Y_v^s and Z_v^s possess this property. But this conjecture seems very plausible; if it is true (9.4) reduces to the Smoluchowski formula (1.1) for b very large.

Formulae (9.1) and (9.4) are so complicated that it is difficult to appreciate all their implications at a glance. The following observations help to display some of their chief features.

1. The importance of the correcting terms increases with the ionic valencies. Also the relaxation terms Z^\dagger and Z^s are greater the smaller the ionic mobilities. This follows because the ω_i enter the q_v^* as the inverse power.

2. A natural question to ask is, for what values of the double layer thickness does the correction to Henry's formula have its greatest value? From (9.1), calculation shows that if Q is taken as constant as b is varied, then the correction is at its maximum for $b \simeq 1.0$. On the other hand, if ζ is taken as constant, the maximum correction arises when $b \simeq 6.0$.

3. Another question of interest is the magnitude of the correction. This depends very much on the nature of the electrolyte owing to the factors q_v and q_v^* . In figure 3 has been plotted the quantity $6\pi\eta eU/EekT$ as a function of $e\zeta/kT$. For one curve U is calculated from (1.2) and for the others, from (9.4). The electrolyte is pure water and for b the value 6.3 has been chosen; this corresponds to the position of maximum correction to Henry's formula. It will be observed that the correction is quite large for $e\zeta/kT > 2$. The effect of the term of order ζ^3 in (9.4) becomes appreciable for $e\zeta/kT > 1.5$, and of the term of order ζ^4 when $e\zeta/kT > 4.0$; at 25° C this corresponds to a value of 100 mV for the zeta potential.

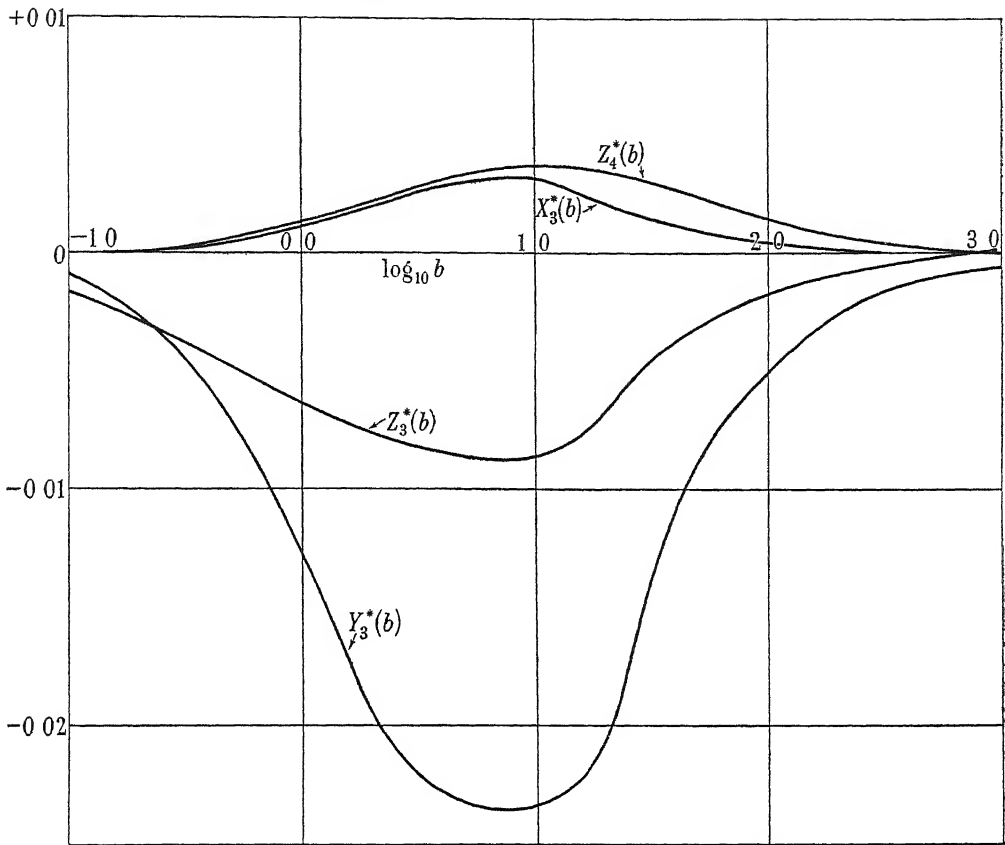


FIGURE 2. Functions $X_3^*(b)$, $Y_3^*(b)$, $Z_3^*(b)$ and $Z_4^*(b)$.

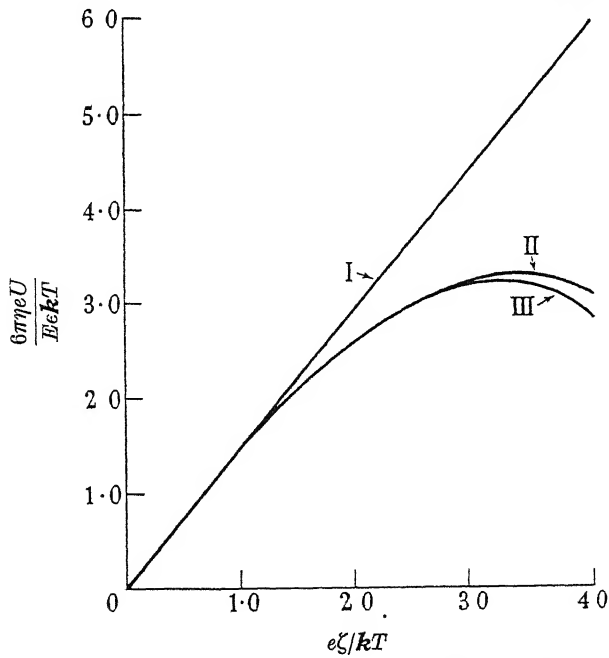


FIGURE 3. Curve (I), equation (1.2). Curve (II), equation (9.4), neglecting terms of order ζ^4 . Curve (III), equation (9.4), neglecting terms of order ζ^5 .

Since zeta potentials sometimes appear to reach values of 150 mV, this result has an important bearing on the use of cataphoresis measurements to find zeta potentials. It shows that unless the double layer thickness is less than a hundredth of, or greater than the radius, the values of ζ calculated from Henry's formula may be completely fallacious unless the zeta potential is quite small. In fact the correction seems so large that some information on terms of higher orders in (9.1) and (9.4) would appear to be very desirable. But the present calculations seem to be as far as one can go by exact methods.

10. DISCUSSION OF PREVIOUS WORK

In view of the divergence between the results of the present paper and those of some other writers, a critical summary of previous work will now be given.

Hermans (1938) has calculated the cataphoretic velocity to the third power of ζ for the limiting case of b large—that is, for a thin double layer. In our notation he finds

$$U = \frac{\epsilon E \zeta}{4\pi\eta} \left(1 + \frac{1}{2} q_2 \frac{e\zeta}{kT} - \frac{5}{72} q_3^* \left(\frac{e\zeta}{kT} \right)^2 \right) \quad b \gg 1. \quad (10.1)$$

This does not agree with our conclusions since we have found that all the correction terms to the Smoluchowski formula vanish when $b \gg 1$. But Hermans's analysis seems open to criticism on several grounds. In the first place the contributions of order ζ^2 and ζ^3 to U , due to the spherically symmetrical part of the electrostatic field are ignored. Also in the final evaluation of U , the electrostatic forces due to the field distortion are taken into account, but the equally important modifications of the fluid stresses by the distortion are ignored.

In another memoir (Komagata 1935) the effect of relaxation on U in a symmetrical electrolyte, there being no restriction on the double layer thickness, has been examined. But the results are invalid for several reasons. To determine the distortion of the ionic atmosphere by the fluid flow, the disturbing effect of the solid particle on the fluid flow is ignored. This means that the results are only valid for thick double layers or for $b \ll 1$. Also the expression for the disturbance of the external field due to the particle is incorrect, since it is made to depend upon the dielectric constants, not the conductivities of the sphere and the electrolyte. Finally the effect of the field distortion on the fluid stresses is taken into account, but the electrostatic forces arising from this distortion are not added.

Finally we come to an elaborate analysis due to Overbeek. In this U is calculated to the third power of ζ , for both symmetrical and unsymmetrical electrolytes. The results are in good general agreement with the present calculations although the method is rather different.

Another paper by Overbeek (1945) gives a simple discussion of the correction to U when b is large. The result (in our notation) that $Z_3(b)$ is of order b^{-1} for b large is confirmed.

I am indebted to Imperial Chemical Industries Ltd. for a research grant under which this work was carried out.

REFERENCES

- Helmholtz, H. v. 1879 *Ann. Phys., Lpz.*, **7**, 337.
 Henry, D. C. 1931 *Proc. Roy. Soc. A*, **133**, 106.
 Hermans, J. J. 1938 *Phil. Mag.* (7), **26**, 650.
 Gronwall, T. H., La Mer, V. K. & Sandved, R. 1928 *Phys. Z.* **28**, 358.
 Komagata, S. 1935 *Res. electrotech. Lab. Minist. Commun. Japan*, no. 387.
 Overbeek, J. Th. G. 1943 *Kolloidbeihfte*, **54**, 287.
 Overbeek, J. Th. G. 1945 *Philips research reports*, **1**, 315.
 Smoluchowski, M. v. 1918 *Z. phys. Chem.* **93**, 129.

The electroviscous effect for suspensions of solid spherical particles

By F. BOOTH, *H. H. Wills Physical Laboratory, Bristol University**

(Communicated by N. F. Mott, *F.R.S.*—Received 19 June 1948—Publication delayed at author's request in order to follow preceding paper)

In this paper a theoretical analysis is made of the electrokinetic phenomenon known as the 'electroviscous effect'. A general formula is given for the effective viscosity of a suspension of solid, spherical, charged non-conducting particles in an electrolyte.

The increase of the effective viscosity due to the surface charge and the ionic double layer surrounding the particles is determined by a modification of Einstein's method for the calculation of the viscosity of solid suspensions. The effective viscosity may be expressed in the form

$$\eta = \eta_0 \left\{ 1 + 2.5(v/V) \left(1 + \sum_{r=1}^{\infty} a_r Q^r \right) \right\},$$

where η_0 is the viscosity of the electrolyte, v the volume of suspension in volume V of solution and Qe is the charge on each particle. It is shown that $a_1 = 0$ and a_2 is determined explicitly.

It is found that the electroviscous contribution to η , for a given charge Q , tends to increase as the thickness of the double layer increases. When the thickness of the double layer is small compared with the radius of the particle the effect vanishes.

A comparison with previous theoretical work is made, and it is shown that much improved agreement with experiment is obtained.

1. INTRODUCTION

The first theoretical investigation of the viscosity of solid suspensions in a fluid was made by Einstein (1906). He showed that the effective viscosity of a suspension of solid spheres in a liquid of viscosity η_0 is given by the formula

$$\eta = \eta_0 \left\{ 1 + \frac{5v}{2V} + O\left(\frac{v}{V}\right)^2 \right\}, \quad (1.1)$$

where v is the volume of the solid, V the total volume of the mixture and v is small compared with V . The expression $O(v/V)^n$ denotes terms of the order $(v/V)^n$. Guth & Simha (1936) added a third term to Einstein's formula and gave

$$\eta = \eta_0 \left\{ 1 + \frac{5v}{2V} + \frac{109}{14} \left(\frac{v}{V}\right)^2 + O\left(\frac{v}{V}\right)^3 \right\}. \quad (1.2)$$

* Present address: King's College, London.

Jeffery (1922) extended Einstein's work to ellipsoidal particles; in this case the factor $\frac{5}{2}$ is replaced by a coefficient depending upon the axial ratio of the particles and their orientation with respect to the flow.

Smoluchowski (1916) pointed out that for a charged particle in an electrolyte the electrical double layer round the particle might be expected to increase the effective viscosity of the suspension. He obtained the relation

$$\eta = \eta_0 \left[1 + \frac{5v}{2V} \left\{ 1 + \frac{1}{\sigma \eta_0 a^2} \left(\frac{\zeta \epsilon}{2\pi} \right)^2 \right\} \right], \quad (1.3)$$

where σ is the specific conductivity of the electrolyte, a the radius of the solid particles, ϵ the dielectric constant of water, and ζ the electrokinetic potential of the particles with respect to the electrolyte. The formula only applies when the thickness of the double layer is small compared with the particle radius.

Smoluchowski gave no proof of relation (1.3), but later Krasny-Ergen (1936) published a derivation; he obtained a formula almost identical with (1.3), the only modification being a factor $\frac{3}{2}$ multiplying the correcting term due to the electrokinetic potential.

Experimental work has confirmed the qualitative result that the presence of the charge on the spherical particle increases the viscosity—a phenomenon known as the 'electroviscous effect'. But quantitative agreement is not very satisfactory, formula (1.3) predicts in general much too great an increase in the viscosity. But the evidence is rather inconclusive because of experimental difficulties.

In the following paper we shall attempt to recalculate the magnitude of the 'electroviscous effect', or modification of the effective viscosity due to the electrical double layer. We shall obtain η without Smoluchowski's restriction that the double layer is thin compared with the size of the solid particle. It will be shown that Smoluchowski's formula requires considerable modification.

The method we shall adopt gives the viscosity in the form

$$\eta = \eta_0 \left[1 + \frac{5v}{2V} \left\{ 1 + \sum_1^{\infty} a_n Q^n \right\} \right], \quad (1.4)$$

where Qe denotes the charge on the particles and e is the electronic charge. In theory the series of coefficients a_1, a_2, \dots can all be calculated, but in practice the method is prohibitively complicated except for the first few coefficients. In this paper we shall only derive a_1 and a_2 , but it is hoped to examine some of the higher coefficients in a later publication.

Since Q can be expressed as a power series in ζ , the electrokinetic potential, an alternative method of writing (1.4) is

$$\eta = \eta_0 \left[1 + \frac{5v}{2V} \left\{ 1 + \sum_1^{\infty} b_n \left(\frac{e\zeta}{kT} \right)^n \right\} \right]. \quad (1.5)$$

2. BASIC ASSUMPTIONS AND GENERAL METHOD OF CALCULATION

To solve the purely hydrodynamical problem treated by Einstein we require only one set of equations, the usual equations for the fluid motion; but in our case we have two sets, one for the motion of the electrolyte, the other for the motion of the ions.

For the electrolyte we make the usual assumptions. We neglect the 'inertia terms' in the equations of motion and assume that there is no turbulence. Also, the electrolyte is regarded as an incompressible fluid, and there is no 'slip' at the surface of the spherical particles. With these assumptions the equations of motion and continuity are respectively

$$\text{grad } p + \rho \text{ grad } \psi = \eta_0 \Delta \mathbf{v}, \quad (2.1)$$

$$\text{div } \mathbf{v} = 0, \quad (2.2)$$

where p is the pressure, \mathbf{v} the fluid velocity, ρ the charge density and ψ the electrostatic potential. The derivation of (2.1) and (2.2) can be found in the preceding paper by the writer on electrophoresis (1950) which will be denoted hereafter by I.

ψ and ρ are related by Poisson's equation

$$\Delta \psi = -4\pi\rho/\epsilon, \quad (2.3)$$

ϵ being the dielectric constant of the electrolyte.

As in Einstein's calculation we shall assume that the volume of the solid is small compared with the total volume of the suspension or that

$$v \ll V. \quad (2.4)$$

In I it is shown that the equations of motion of ions are given by the equations

$$\bar{\mathbf{u}}_i = \mathbf{v} - w_i \left(\frac{kT}{m_i} \text{grad } m_i + ez_i \text{grad } \psi \right), \quad (2.5)$$

$$\text{div } (m_i \bar{\mathbf{u}}_i) = 0, \quad (2.6)$$

where i distinguishes the type of ion and $\bar{\mathbf{u}}_i$, w_i , m_i and z_i are the mean velocity, mobility, concentration and valency respectively of ions of type i .

We assume that all the quantities ϵ , η_0 and w_i are uniform throughout the electrolyte.

We also require some assumptions on the conditions at the solid-electrolyte interface. A full discussion of these is given in I; precisely the same assumptions are made in the present paper, namely:

(1) The thickness of the 'surface phase', or region containing the surface charge on the particle, is small compared with the radius of the particle.

(2) The ions in the surface phase are immobile—they cannot move laterally over the surface.

(3) The surface-charge density at any point is fixed; it remains unchanged when the electrolyte is set in motion.

(4) The potential across the surface phase is also unaffected by the motion of the electrolyte.

(5) All the basic assumptions so far made were also made in I. In the present paper a further condition will be added, that the double layer is small in thickness compared with the average distance between the neighbouring particles. Now the quantity χ^{-1} defined in § 4 is a measure of the thickness of the double layer; hence we postulate that

$$\chi^{-1} \ll r_0, \quad \text{where} \quad 4\pi r_0^3 N/3 = V, \quad (2.7)$$

where N is the total number of particles in the suspension.

Since we take r_0 to be much larger than the radius of the solid sphere, or a , in view of (2.4), condition (2.7) imposes no restriction on the relation of the double-layer thickness to the size of the particle. If condition (2.7) is not imposed it becomes necessary to consider the interaction of the charges and double layers of neighbouring particles.

Finally, we shall consider only non-conducting solid particles.

This concludes our review of the basic assumptions, and we shall now sketch our method of calculation. To calculate the effective viscosity from equations (2.1) to (2.6) we shall employ a method devised by Fröhlich & Sack (1946) for the analysis of the rheological properties of dispersions. We shall refer to their paper hereafter by the symbol II. Our problem is to determine the coefficient of viscosity η which a volume of *uniform* fluid would have if it exhibited the same stress-strain relationships on its surface as a volume of the suspension sufficiently large to contain many solid particles. For convenience we may call the fictitious fluid of uniform texture which is equivalent to the suspension, the 'macroscopic fluid'.

Suppose we apply a system of stresses to a sphere of radius R of the macroscopic fluid so that the velocity distribution in the sphere has rectangular components

$$u_i = \sum_{j=1}^3 \alpha_{ij} x_j \quad (i = 1, 2, 3), \quad (2.8)$$

where x_i ($i = 1, 2, 3$) denotes rectangular co-ordinates. Equation (2.8) will satisfy (2.2) and also represent irrotational motion, provided that

$$\sum_i \alpha_{ii} = 0, \quad \alpha_{ij} = \alpha_{ji} \quad (i \neq j). \quad (2.9)$$

The stress components required to give the flow (2.8) are given by

$$S_{ij} = -p\delta_{ij} + \eta \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) = -p\delta_{ij} + 2\eta\alpha_{ij}, \quad (2.10)$$

where p is the pressure, and $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$.

Suppose we now replace a sphere of the macroscopic fluid, of radius r_0 and concentric with the sphere of radius R , by its 'microscopic structure', that is a solid particle of radius a at its centre, surrounded by a spherical shell of fluid of viscosity η_0 between radii $r = a$ and $r = r_0$. If the system of stresses (2.10) is applied to the surface $r = R$ of this second model, the flow will be different from that of the macroscopic fluid. But if we make R sufficiently large the flow on $r = R$ for the second model will be given by equations (2.8). From this condition and the boundary conditions at $r = a$ and $r = r_0$ we can calculate the viscosity of the macroscopic fluid. The boundary conditions to be satisfied are continuity of velocity at $r = a$ and $r = r_0$, and stress continuity at $r = r_0$.

The only essential difference between our present calculation and that of Fröhlich & Sack in § A of II is that the flow in the spherical shell between $r = a$ and $r = r_0$ is modified by the presence of the charge on the solid and its double layer. For this region we must solve (2.1) and (2.2) for \mathbf{v} and p , utilizing equations (2.3), (2.5) and (2.6) to find ρ and ψ .

For the purpose of solving equations (2.3) and (2.5) it is convenient to write \mathbf{v} , p and ψ as the sum of two velocities, pressures, and potentials respectively,

$$\mathbf{v} = \mathbf{v}^* + \mathbf{v}^\dagger, \quad (2.11)$$

$$p = p^* + p^\dagger \quad (2.12)$$

and
$$\psi = \psi_1 + \phi. \quad (2.13)$$

\mathbf{v}^* and p^* denote the solutions for \mathbf{v} and p if the charge on the sphere were zero. \mathbf{v}^\dagger and p^\dagger denote the disturbances in \mathbf{v} and p brought about by the presence of the charge on the particle.

A rather similar scheme obtains for ψ . ψ_1 denotes the solution of (2.3), (2.5) and (2.6), with \mathbf{v} set equal to zero; it is the ordinary solution for the potential when the liquid is at rest. When the electrolyte is moving the ions tend to be dragged along with the liquid, so modifying the field due to the double layer. Thus ϕ represents the distortion of ψ_1 , due to the fluid motion. For our purpose we shall only eventually require \mathbf{v}^* and \mathbf{v}^\dagger , but a knowledge of ϕ is necessary to find \mathbf{v}^\dagger .

In the next section we shall consider briefly the velocity \mathbf{v}^* . §4 contains the relevant solutions for ψ_1 . Using these preliminary results we shall calculate in §5 both ϕ and \mathbf{v}^\dagger to the required degree of accuracy. Finally in §6 we complete the calculation of the effective viscosity by using the solutions for ϕ and \mathbf{v}^\dagger . §7 is devoted to a discussion of the results and a brief comparison with the experimental data.

3. CALCULATION OF THE VELOCITY DISTRIBUTION \mathbf{v}^* (NO CHARGE ON SOLID PARTICLES)

The relevant solutions may be found in §A of II; we shall therefore merely set down those conclusions which we shall require.

Transforming to spherical polar co-ordinates

$$(x = r \sin \theta \cos \alpha, \quad y = r \sin \theta \sin \alpha, \quad z = r \cos \theta),$$

we can write the stress components in the form

$$S_{rr} = -p + 2\eta_0 f, \quad S_{r\theta} = \eta_0 \frac{\partial f}{\partial \theta}, \quad S_{r\alpha} = \frac{\eta_0}{\sin \theta} \frac{\partial f}{\partial \alpha}, \quad (3.1)$$

$$f(\theta, \alpha) = \alpha_{11} \sin^2 \theta \cos^2 \alpha + \alpha_{22} \sin^2 \theta \sin^2 \alpha + \alpha_{33} \cos^2 \theta \\ + 2\alpha_{23} \sin \theta \cos \theta \sin \alpha + 2\alpha_{13} \sin \theta \cos \theta \cos \alpha + 2\alpha_{12} \sin^2 \theta \cos \alpha \sin \alpha, \quad (3.2)$$

where f is a spherical harmonic of the second degree.

We need the velocity of the fluid in the region $a < r < r_0$ when the stress system (3.1) is applied to the surface $r = R$ and $\rho = \psi = 0$. Eventually it will be found that we only require the r component of \mathbf{v} , and this is easily determined from equations (23), (35) and (36) of paper II. We have

$$v_r^* = r f \left[1 - \frac{5a^3}{2r^3} + \frac{3a^5}{2r^5} + O\left(\frac{a^3}{r^3}\right) \right]. \quad (3.3)$$

This concludes our determination of \mathbf{v}^* and we shall now proceed to find ψ_1 .

4. THE UNDISTORTED POTENTIAL

The solution for the potential when the fluid velocity is everywhere zero has been given by the writer in paper I. We here state the conclusions only.

We found that ψ_1 could be expressed as a power series in Q , where Qe is the charge on the spherical particle,

$$\psi_1 = \sum_{m=1}^{\infty} e^{2m-1} (ea)^{-m} (-kT)^{1-m} Q^m \lambda_m (\chi r). \quad (4.1)$$

Since we shall only ultimately require the first term of series (4.1) we merely give $\lambda_1(x)$. We have

$$\lambda_1(x) = \frac{b}{1+b} \frac{e^{b-x}}{x}, \quad (4.2)$$

where $b = \chi a$ and $\chi^2 = 4\pi e^2 \sum_{i=1}^s n_i z_i^2 / \epsilon kT$, and n_i is the concentration of the ions of type i at a large distance from the particle. e is the electronic charge, k Boltzmann's constant and T the absolute temperature.

5. CALCULATION OF \mathbf{v} AND p .

Having obtained \mathbf{v}^* and ψ_1 , we can now proceed to obtain the general solution for the velocity \mathbf{v} . Let m_{1i} denote the ionic concentrations when the fluid velocity is everywhere zero. Then by Boltzmann's relation we have

$$m_{1i} = n_i \exp(-ez_i \psi_1 / kT). \quad (5.1)$$

Then if we write

$$m_i = m_{1i} + s_i, \quad (5.2)$$

s_i can be regarded as the 'distortion' of the ionic concentrations due to the fluid motion. We have

$$\Delta\phi = -\frac{4\pi e}{\epsilon} \sum_{i=1}^s z_i s_i. \quad (5.3)$$

Multiplying equation (2.5) by m_i and taking its divergence, we obtain, after some simplification, the following differential equation relating p_i , ϕ and \mathbf{v} :

$$\begin{aligned} kT \Delta s_i + ez_i \left[s_i (\Delta\phi + \Delta\psi_1) + \frac{d\psi_1}{dr} \frac{\partial s_i}{\partial r} + (\text{grad } s_i \cdot \text{grad } \phi) \right. \\ \left. + n_i \exp\left(-\frac{ez_i \psi_1}{kT}\right) \left(\Delta\phi - \frac{ez_i}{kT} \frac{d\psi_1}{dr} \frac{\partial \phi}{\partial r} \right) \right] \\ = \frac{1}{w_i} \text{div } \mathbf{v} \left\{ n_i \exp\left(-\frac{ez_i \psi_1}{kT}\right) + s_i \right\}. \end{aligned} \quad (5.4)$$

Now for small values of \mathbf{v} we should expect s_i and ϕ to vary as some linear function of the velocity and possibly of its derivatives. Hence three of the terms of equation (5.4), namely $\Delta\phi$, $(\text{grad } s_i \cdot \text{grad } \phi)$ and $\text{div } \mathbf{v} s_i$, fall into the same category as the 'inertia terms' in the equations of motion. Since we have neglected these terms, we also drop the similar trio in (5.4).

To solve equation (5.4) we assume developments for s_i , ϕ and v_r^\dagger in series of powers of Q , Qe being the total charge on the sphere when the system is at rest. Let

$$\phi = f(\theta, \alpha) \sum_{\nu=1}^{\infty} e^{2\nu-1} (\epsilon a)^{-\nu} (kT)^{1-\nu} Q^\nu \phi_\nu(r), \quad (5.5)$$

$$s_i = f(\theta, \alpha) \sum_{\nu=1}^{\infty} e^{2\nu-1} (\epsilon a)^{-\nu} (kT)^{1-\nu} Q^\nu \gamma_{i,\nu}(r), \quad (5.6)$$

$$v_r^\dagger = f(\theta, \alpha) \sum_{\nu=1}^{\infty} e^{2\nu-2} \eta_0^{-1} (\epsilon a)^{1-\nu} (kT)^{2-\nu} Q^\nu \beta_{r,\nu}(r). \quad (5.7)$$

v_r^\dagger , v_θ^\dagger and v_α^\dagger denote the components of \mathbf{v}^\dagger in the r , θ and α directions. It is also convenient to have similar developments for v_θ^\dagger , v_α^\dagger and p^\dagger ,

$$v_\theta^\dagger = \frac{\partial f}{\partial \theta} \sum_{\nu=1}^{\infty} e^{2\nu-2} \eta_0^{-1} \epsilon^{1-\nu} a^{-1-\nu} (kT)^{2-\nu} Q^\nu \beta_{\theta,\nu}(r), \quad (5.8)$$

and a similar expression for v_α^\dagger with $\partial f / \partial \theta$ replaced by $(1/\sin \theta)(\partial f / \partial \alpha)$ and $\beta_{\theta,\nu}$ by $\beta_{\alpha,\nu}$. We also set

$$p^\dagger = f \sum_{\nu=1}^{\infty} e^{2\nu-2} \epsilon^{1-\nu} a^{-2-\nu} (kT)^{2-\nu} Q^\nu \pi_\nu(r). \quad (5.9)$$

Here f is defined by equation (3.2).

In the above expansions all the expressions $f\phi_\nu$, $f\beta_{r,\nu}$, $f\beta_{\theta,\nu}$, $f\beta_{\alpha,\nu}$, and $f\pi_\nu$ will be dimensionless functions of r . Strictly we should use for ϕ , etc., perfectly general expansions in terms of harmonics of all degrees from zero upwards. It can be shown, however, by methods very similar to those developed in I in § 5, that all portions of such an expansion will vanish except the portion multiplying the harmonic $f(\theta, \alpha)$. This conclusion follows from the third and fourth assumptions we have made on the nature of the solid-liquid interface. We have accordingly written ϕ , s_i and \mathbf{v}^\dagger in the above way.

Substituting for ψ_1 from (4.1) and using (5.5) to (5.9) in (5.4) we get a series of equations for the successive functions $\phi_1, \phi_2, \dots, \gamma_{1,1}, \gamma_{1,2}, \dots, \gamma_{2,1}, \gamma_{2,2}, \dots, \gamma_{s,1}, \gamma_{s,2}, \dots$ and $\beta_{r,1}, \beta_{r,2}, \dots$

$$\left. \begin{aligned} \Delta_2^{(r)}(kT\gamma_{i,1} + ez_i n_i \phi_1) &= -\frac{n_i ez_i}{w_i kT} V(r) \frac{d\lambda_1}{dr} \quad (i = 1, 2, \dots, s), \\ \Delta_2^{(r)}(kT\gamma_{i,2} + ez_i n_i \phi_2) &= -kT z_i \left(\gamma_{i,1} \Delta_0^{(r)} \lambda_1 + \frac{d\lambda_1}{dr} \frac{d\gamma_{i,1}}{dr} \right) + n_i ez_i^2 \left(\lambda_1 \Delta_2^{(r)} \phi_1 + \frac{d\lambda_1}{dr} \frac{d\phi_1}{dr} \right) \\ &\quad + \frac{n_i ez_i}{w_i kT} V(r) \left(z_i \lambda_1 \frac{d\lambda_1}{dr} - \frac{d\lambda_2}{dr} \right) - \frac{n_i ez_i kT}{\eta_0 w_i a e} \frac{d\lambda_1}{dr} \beta_{r,1} \\ \Delta_2^{(r)}(kT\gamma_{i,n} + ez_i n_i \phi_n) &= \dots \end{aligned} \right\} \quad (i = 1, 2, \dots, s), \quad (5.10)$$

$$\text{where} \quad \Delta_n^{(r)} = \frac{d^2}{dr^2} + \frac{2}{r} \frac{d}{dr} - \frac{n(n+1)}{r^2}, \quad (5.11)$$

$$\text{and we write} \quad v_r^* = V(r) f(\theta, \alpha). \quad (5.12)$$

From equation (3.3) it follows that

$$V(r) = r - \frac{5a^3}{2r^2} + \frac{3a^5}{2r^4} + O\left(\frac{a}{r_0}\right)^3. \quad (5.13)$$

To determine the effective viscosity we have to evaluate the successive functions $\beta_{r,1}$, $\beta_{\theta,1}$, $\beta_{\alpha,1}$, $\beta_{r,2}$, $\beta_{\theta,2}$, $\beta_{\alpha,2}$, ... These can be found from equations (5.10) and equations (2.1) and (2.2) for the motion of the electrolyte by the following method:

(1) The first of equations (5.10) is solved for the unknowns $\gamma_{i,1}$ and ϕ_1 .
 (2) Using these solutions we can find all the six functions $\beta_{r,1}$, ..., $\beta_{\alpha,2}$ from equations (2.1) and (2.2).

(3) In the second of (5.10), all quantities are now known except $\gamma_{i,2}$ and ϕ_2 . Integrating, and using the boundary conditions, we can find these quantities. The same applies to the third of equations (5.10) which gives $\gamma_{i,3}$ and ϕ_3 .

(4) To solve the fourth of equations (5.10) for $\gamma_{i,4}$ and ϕ_4 we require $\beta_{r,3}$. This must be found from the equations of motion, using the solutions for ϕ_1 , ϕ_2 and ϕ_3 .

It is evident that this process of alternately solving the appropriate equation of set (5.10) and the equations of motion can, in theory, be continued indefinitely, and the velocity found correct to any desired power of Q , the particle charge. But, in practice, the solution becomes very complicated after the first few stages. In the present paper we shall evaluate the effective viscosity correct to Q^2 and therefore we evaluate $\beta_{r,1}$, $\beta_{\theta,1}$, $\beta_{\alpha,1}$, $\beta_{r,2}$, $\beta_{\theta,2}$, $\beta_{\alpha,2}$, only; thus we carry out stages 1 and 2 only of the above programme.

The first step is the integration of (5.10) for $\gamma_{i,1}$ and ϕ_1 . We have

$$kT\gamma_{i,1} + n_i e z_i \phi_1 = A_{i,1} r^2 + \frac{\beta_{i,1}}{r^3} + \frac{n_i e z_i}{5w_i kT} \int_r^\infty V(z) \frac{d\lambda_1(\chi z)}{dz} z^2 \left(\frac{r^2}{z^3} - \frac{z^2}{r^3} \right) dz, \quad (5.14)$$

where $A_{i,1}$ and $B_{i,1}$ are constants. Since $\gamma_{i,1}$ and ϕ_1 both tend to zero at infinity, $A_{i,1} = 0$. From (4.2) and (5.12) the integral can be evaluated and we can write

$$kT\gamma_{i,1} + n_i e z_i \phi_1 = \frac{B_{i,1}}{r^3} + \frac{n_i e z_i}{\chi^2 w_i kT} F_1(b, \chi r), \quad (5.15)$$

where

$$F_1(b, x) = \frac{b e^{b-x}}{5(1+b)} \left[5 \left(1 + \frac{3}{x} + \frac{6}{x^2} + \frac{6}{x^3} \right) - \frac{5b^3}{2} \left(-\frac{x}{8} + \frac{1}{8} - \frac{1}{4x} + \frac{3}{4x^2} + \frac{2}{x^3} + \frac{x^2 \text{Ei}(x) e^x}{8} \right) \right. \\ \left. + \frac{3}{2} b^5 \left(-\frac{x}{144} + \frac{1}{144} - \frac{1}{72x} + \frac{1}{24x^2} - \frac{1}{6x^3} + \frac{5}{6x^4} + \frac{x^2 \text{Ei}(x) e^x}{144} \right) \right]. \quad (5.16)$$

To find $B_{i,1}$ we must consider the boundary conditions at the surface. We have assumed no lateral movement of ions in the surface phase, and therefore it follows that the net flow of each type of ion towards the surface vanishes at all points of the surface. Hence if $\bar{\mathbf{u}}_{i,r}(r, \theta, \alpha)$ denotes the radial component of the average ionic velocity $\bar{\mathbf{u}}_i$, then

$$\bar{\mathbf{u}}_{i,r}(a, \theta, \alpha) = 0 \quad \text{for all } \theta \text{ and } \alpha. \quad (5.17)$$

From (2.7), since $v_r = 0$ at $r = a$, this becomes

$$kT \left. \frac{\partial m_i}{\partial r} \right|_a + m_i e z_i \left. \frac{\partial \psi}{\partial r} \right|_a = 0,$$

$$\text{or, from (5.5) and (5.6),} \quad kT \left. \frac{d\gamma_{i,1}}{dr} \right|_a + n_i e z_i \left. \frac{d\phi_1}{dr} \right|_a = 0. \quad (5.18)$$

Applying this condition to (5.15) we find

$$B_{i,1} = \frac{n_i e z_i a^4}{3 \chi w_i k T} F_1'(b, b), \quad (5.19)$$

where

$$F_1'(b, b) = \left. \frac{dF(b, x)}{dx} \right|_b. \quad (5.20)$$

From Poisson's relation we have

$$\Delta_2^{(s)} \phi_1 = -\frac{4\pi e}{\epsilon} \sum_{i=1}^s z_i \gamma_{i,1}. \quad (5.21)$$

Hence, multiplying (5.15) by $4\pi e z_i / \epsilon k T$ and summing over i from 1 to s , we obtain a differential equation for ϕ_1 only,

$$(\Delta_2^{(r)} - \chi^2) \phi_1(r) = -\frac{4\pi e^2}{\epsilon \chi^2 k^2 T^2} \sum_{i=1}^s \frac{n_i z_i^2}{w_i} \left[F_1(b, \chi r) + \frac{\chi a^4}{3r^3} F_1'(b, b) \right]. \quad (5.22)$$

ϕ_1 could be found by integrating this equation once more, but since we do not eventually require ϕ_1 itself but only the quantity $(\Delta_2^{(r)} - \chi^2) \phi_1$, we shall not proceed further with the integration of (5.22)

This completes the first stage of our calculation, and we now proceed to the second stage—the determination of $\beta_{r,1}, \dots, \beta_{\alpha,2}$ from the equations of motion and continuity of the electrolyte.

We evidently require the solution of equations (2.1) and (2.2) where ψ is of the form

$$\psi = \psi_1 + \Psi f, \quad (5.23)$$

and Ψ is a function of r only.

From Poisson's relation

$$\rho = -\frac{\epsilon}{4\pi} (\Delta_0^{(r)} \psi_1 + f \Delta_2^{(r)} \Psi). \quad (5.24)$$

Hence if we drop the term involving products of Ψ (since this is of the same order as the 'inertia terms') (2.1) becomes

$$\eta_0 \text{curl curl } \mathbf{v} + \text{grad } p' = \frac{\epsilon}{4\pi} [f \Delta_2^{(r)} \Psi \text{grad } \psi_1 + \Delta_0^{(r)} \psi_1 \text{grad } f \Psi], \quad (5.25)$$

where

$$p' = p + \frac{\epsilon}{4\pi} \int_r^\infty \Delta_0^{(r)} \psi_1 \frac{d\psi_1}{dr} dr. \quad (5.26)$$

Taking the curl of (5.25) we have

$$\eta_0 \text{curl curl curl } \mathbf{v} = \frac{\epsilon}{4\pi} [[\text{grad } (f \Delta_2^{(r)} \Psi) \wedge \text{grad } \psi_1] + [\text{grad } \Delta_0^{(r)} \psi_1 \wedge \text{grad } f \Psi]]. \quad (5.27)$$

If we set $\mathbf{w} = \text{curl } \mathbf{v}$, then this equation gives for the three components of $\text{curl curl } \mathbf{w}$

$$\left. \begin{aligned} \eta_0 \text{curl curl}_r \mathbf{w} &= 0, \\ \eta_0 \text{curl curl}_\theta \mathbf{w} &= \frac{\epsilon}{4\pi r \sin \theta} \frac{\partial f}{\partial \alpha} \left(\frac{d\psi_1}{dr} \Delta_2^{(r)} \Psi - \Psi \frac{d}{dr} (\Delta_0^{(r)} \psi_1) \right), \\ \eta_0 \text{curl curl}_\alpha \mathbf{w} &= -\frac{\epsilon}{4\pi r} \frac{\partial f}{\partial \theta} \left(\frac{d\psi_1}{dr} \Delta_2^{(r)} \Psi - \Psi \frac{d}{dr} (\Delta_0^{(r)} \psi_1) \right). \end{aligned} \right\} \quad (5.28)$$

To solve these equations we first write the components of $\text{curl curl } \mathbf{w}$ in terms of the components of \mathbf{w} . This is easily accomplished by the usual formulae for the differential operators. We find, for example,

$$\text{curl curl}_\alpha \mathbf{w} = \frac{1}{r \sin \theta} \frac{\partial^2 w_r}{\partial r \partial \alpha} - \frac{1}{r} \frac{\partial^2}{\partial r^2} (r w_\alpha) - \frac{1}{r^2} \frac{\partial}{\partial \theta} \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} (w_\alpha \sin \theta) + \frac{1}{r^2} \frac{\partial}{\partial \theta} \frac{1}{\sin \theta} \frac{\partial w_\theta}{\partial \alpha}. \quad (5.29)$$

If now we try a solution of (5.28) in the form

$$w_r = 0, \quad w_\theta = -\frac{W(r)}{\sin \theta} \frac{\partial f}{\partial \alpha}, \quad w_\alpha = W(r) \frac{\partial f}{\partial \theta}, \quad (5.30)$$

where $W(r)$ is a function of r but not of θ or α , we find that the expressions (5.30) form a solution provided $W(r)$ satisfies the equation

$$\Delta_2^{(s)} W = \frac{\epsilon}{4\pi\eta_0 r} \left(\frac{d\psi_1}{dr} \Delta_2^{(s)} \Psi - \Psi \frac{d}{dr} \Delta_0^{(s)} \psi_1 \right). \quad (5.31)$$

The solution of this equation is easily determined by the method of variation of the constants

$$W = Ar^2 + \frac{B}{r^3} + \frac{\epsilon}{20\pi\eta_0} \int_r^\infty \left(\frac{z^3}{r^3} - \frac{r^2}{z^2} \right) \left(\frac{d\psi_1(\chi z)}{dz} \Delta_2^{(s)} \Psi(z) - \Psi(z) \frac{d}{dz} (\Delta_0^{(s)} \psi_1(\chi z)) \right) dz, \quad (5.32)$$

where A and B are constants.

Having determined \mathbf{w} we must now find \mathbf{v} from the two equations

$$\mathbf{w} = \text{curl } \mathbf{v}, \quad (5.33)$$

$$\text{div } \mathbf{v} = 0. \quad (5.34)$$

We assume a solution in the form

$$v_r = V_1 f, \quad v_\theta = \frac{V_2}{r} \frac{\partial f}{\partial \theta}, \quad v_\alpha = \frac{V_2}{r \sin \theta} \frac{\partial f}{\partial \alpha}, \quad (5.35)$$

where V_1 and V_2 are functions of r only.

From equations (5.35) we have

$$\text{div } \mathbf{v} = \frac{f}{r^2} \frac{d}{dr} (r^2 V_1) + \frac{V_2}{r^2 \sin \theta} \left\{ \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{\sin \theta} \frac{\partial^2 f}{\partial \alpha^2} \right\} = 0. \quad (5.36)$$

Since
$$\frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial f}{\partial \theta} \right) + \frac{1}{\sin \theta} \frac{\partial^2 f}{\partial \alpha^2} = -6 \sin \theta f, \quad (5.37)$$

it follows from (5.36) that V_1 and V_2 satisfy the equation

$$\frac{d}{dr} r^2 V_1 - 6V_2 = 0. \quad (5.38)$$

Now

$$\text{curl}_r v = \frac{1}{r \sin \theta} \left[\frac{\partial}{\partial \theta} \sin \theta v_\alpha - \frac{\partial v_\theta}{\partial \alpha} \right] = 0.$$

Hence equations (5.35) satisfy the condition $w_r = 0$. Similarly, we find that the second and third of equations (5.30) are satisfied provided that

$$rW = \frac{dV_2}{dr} - V_1. \quad (5.39)$$

Equations (5.38) and (5.39) determine V_1 and V_2 ; eliminating V_1 we find that V_2 satisfies the equation

$$\Delta_2^{(r)} V_2 = r \frac{dW}{dr} + 3W. \quad (5.40)$$

Integrating this equation we have

$$V_2 = Cr^2 + \frac{D}{r^3} - \frac{1}{5} \int^r \left(z \frac{dW(z)}{dz} + 3W(z) \right) \left(\frac{z^4}{r^3} - \frac{r^2}{z} \right) dz, \quad (5.41)$$

where C and D are again arbitrary constants.

From (5.39) we find for V_1 ,

$$V_1 = 2Cr - \frac{3B}{r^4} - rW(r) + \frac{1}{5} \int^r \left(z \frac{dW(z)}{dz} + 3W(z) \right) \left(\frac{3z^4}{r^4} + \frac{2r}{z} \right) dz. \quad (5.42)$$

Collecting our results we find

$$\left. \begin{aligned} v_r &= f \left[\frac{3}{7} Ar^3 - \frac{B}{r^2} + 2Cr - \frac{3D}{r^4} - rU(r) + \frac{dX(r)}{dr} \right], \\ v_\theta &= \frac{\partial f}{\partial \theta} \left[\frac{5}{14} Ar^3 + Cr + \frac{D}{r^4} + \frac{X(r)}{r} \right], \\ v_\alpha &= \frac{1}{\sin \theta} \frac{\partial f}{\partial \alpha} \left[\frac{5}{14} Ar^3 + Cr + \frac{B}{r^4} + \frac{X(r)}{r} \right], \end{aligned} \right\} \quad (5.43)$$

$$\left. \begin{aligned} \text{where} \quad X(r) &= \frac{1}{5} \int_r^\infty \left(z \frac{dU(z)}{dz} + 3U(z) \right) \left(\frac{z^4}{r^3} - \frac{r^2}{z} \right) dz \\ \text{and} \quad U(r) &= \frac{\epsilon}{20\pi\eta_0} \int_r^\infty \left(\frac{d\psi_1(\chi z)}{dz} \Delta_2^{(z)} \Psi - \Psi \frac{d}{dz} \Delta_0^{(z)} \psi_1(\chi z) \right) \left(\frac{z^3}{r^3} - \frac{r^2}{z^2} \right) dz. \end{aligned} \right\} \quad (5.44)$$

Equations (5.43) determine the velocity completely, provided the potential Ψ is known; when $\Psi = 0$ they become equivalent to expressions (23) and (24) of paper II.

Taking the divergence of equation (5.25), we have

$$\Delta p' = \frac{\epsilon f}{4\pi} \left[2\Delta_0^{(r)} \psi_1(\chi r) \Delta_2^{(r)} \Psi(r) + \frac{d}{dr} \Delta_0^{(r)} \psi(\chi r) \frac{d\Psi_1(r)}{dr} + \frac{d\psi_1(\chi r)}{dr} \frac{d}{dr} \Delta_2^{(r)} \Psi(r) \right]. \quad (5.45)$$

We have dropped terms involving products of $\Psi(r)$.

The general solution of (5.45) is

$$\begin{aligned} p' &= f \left[Pr^2 + \frac{Q}{r^3} + \frac{\epsilon}{20\pi} \int_r^\infty \left(2\Delta_0^{(z)}(\chi z) \Delta_2^{(z)} \Psi(z) \right. \right. \\ &\quad \left. \left. + \frac{d}{dz} \Delta_0^{(z)} \psi_1(\chi z) \frac{d\Psi(z)}{dz} + \frac{d\psi_1(\chi z)}{dz} \frac{d}{dz} \Delta_2^{(z)} \Psi(z) \right) \left(\frac{z^4}{r^3} - \frac{r^2}{z} \right) dz. \right] \quad (5.46) \end{aligned}$$

The integration constants P and Q are found immediately by substituting p' as given by (5.46) in equation (5.25)

$$P = 3\eta_0 A, \quad Q = -2\eta_0 B. \quad (5.47)$$

Having obtained equations (5.43) and (5.46) stage 2 of our calculation is now complete, apart from fixing the constants A to D from the boundary conditions. In the present paper, however, we shall not go on to solve for the higher terms.

All the material is now available, therefore, for the evaluation of the viscosity η to the degree of approximation we require.

6. CALCULATION OF THE EFFECTIVE VISCOSITY

From paper II we have for the fluid flow in the region $r < r_0 < R$

$$v_r = f\left(r - \frac{3D_1}{r^4}\right), \quad v_\theta = \frac{\partial f}{\partial \theta}\left(\frac{r}{2} + \frac{D_1}{r^4}\right), \quad v_\alpha = \frac{1}{\sin \theta} \frac{\partial f}{\partial \alpha}\left(\frac{r}{2} + \frac{D_1}{r^4}\right), \quad (6.1)$$

where D_1 is a constant.

From (2.10) we have for the stresses

$$S_{rr} = \eta f\left(2 + \frac{24D_1}{r^5}\right), \quad S_{r\theta} = \eta \frac{\partial f}{\partial \theta}\left(1 - \frac{8D_1}{r^5}\right), \quad S_{r\alpha} = \frac{\eta}{\sin \theta} \frac{\partial f}{\partial \alpha}\left(1 - \frac{8D_1}{r^5}\right). \quad (6.2)$$

If we apply the boundary conditions to equations (5.43) we find at the surface $r = a$

$$\left. \begin{aligned} \frac{3Aa^3}{7} - \frac{B}{a^2} + 2Ca - \frac{3D}{a^4} - aU(a) + \frac{dX}{dr}\bigg|_a &= 0, \\ \frac{5Aa^3}{14} + Ca + \frac{D}{a^4} + \frac{X(a)}{a} &= 0. \end{aligned} \right\} \quad (6.3)$$

We have assumed in (6.3) that on the surface of the particle

$$v_r = v_\theta = v_\alpha = 0. \quad (6.4)$$

These conditions are true for the velocity distribution \mathbf{v}^* , but to apply them to the case when the particle possesses a potential across a double layer requires justification. We show in an appendix, however, that the total force and total couple on the particle both vanish and therefore equations (6.4) and (6.3) are valid.

At the surface $r = r_0$ both the velocity and the stress must be continuous. These conditions give the four equations

$$\left. \begin{aligned} \frac{3Ar_0^3}{7} - \frac{B}{r_0^2} + 2Cr_0 - \frac{3D}{r_0^4} &= r_0 - \frac{3D_1}{r_0^4}, \\ \frac{5Ar_0^3}{14} + Cr_0 + \frac{D}{r_0^4} &= \frac{r_0}{2} + \frac{D_1}{r_0^4}, \end{aligned} \right\} \quad (6.5)$$

and

$$\left. \begin{aligned} \eta_0 \left[-\frac{3Ar_0^2}{7} + \frac{6B}{r_0^3} + 4C + \frac{24D}{r_0^5} \right] &= \eta \left[2 + \frac{24D_1}{r_0^5} \right], \\ \eta_0 \left[\frac{8Ar_0^2}{7} - \frac{B}{r_0^3} + 2C - \frac{8D}{r_0^5} \right] &= \eta \left[1 - \frac{8D_1}{r_0^5} \right]. \end{aligned} \right\} \quad (6.6)$$

Since we have assumed that the double layer does not extend to the sphere $r = r_0$, it follows that

$$X(r_0), \quad \frac{dX}{dr}\bigg|_{r_0}, \quad U(r_0) \quad (6.7)$$

all vanish, and so equations (6.5) and (6.6) do not involve terms depending upon Q .

Solving the six equations (6.3) to (6.6) for the six unknowns A, B, C, D, D_1 and η and remembering that $a \ll r_0$, we find

$$\left. \begin{aligned} \eta &= \eta_0 \left[1 + \frac{a^3}{r_0^3} \left\{ \frac{5}{2} + Y(a) \right\} \right], \\ A &= 0, \quad B = a^3 \left[\frac{5}{2} + Y(a) \right], \quad C = \frac{1}{2} + \frac{a^3}{r_0^3} \left[\frac{1}{2} + \frac{1}{5} Y(a) \right], \\ D &= -\frac{1}{2} a^5 - a^3 X(a), \quad D_1 = a^3 r_0^3 \left[\frac{1}{2} + \frac{1}{5} Y(a) \right], \end{aligned} \right\} \quad (6.8)$$

where
$$Y(r) = \frac{3}{r^2} X(r) - U(r) + \frac{1}{r} \frac{dX}{dr}. \quad (6.9)$$

Substituting from equation (5.22) for Ψ , we find that η can be written in the form

$$\eta = \eta_0 \left[1 + \frac{5v}{2V} \left\{ 1 + \frac{2e^2 Q^2}{5\eta_0 \epsilon a^4} \left(3X_1(a) + a \frac{dX_1(r)}{dr} \Big|_a - a^2 U_1(a) \right) \right\} \right], \quad (6.10)$$

where
$$U_1(r) = \frac{1}{20\pi} \int_r^\infty \frac{d\lambda_1(\lambda z)}{dz} (\Delta_2^{(e)} \phi_1(z) - \chi^2 \phi_1(z)) \left(\frac{z^3}{r^3} - \frac{r^2}{z^2} \right) dz \quad (6.11)$$

and
$$X_1(r) = \frac{1}{5} \int_r^\infty \left(z \frac{dU_1(z)}{dz} + 3U_1(z) \right) \left(\frac{z^4}{r^3} - \frac{r^2}{z} \right) dz. \quad (6.12)$$

From formula (6.10) we see immediately that the correction to the viscosity due to the electroviscous effect is of order Q^2 , or, equivalently, of order ζ^2 . Hence in formulae (1.4) and (1.5) we have
$$a_1 = b_1 = 0. \quad (6.13)$$

The evaluation of the three quantities $U_1(a)$, $X_1(a)$ and $X'_1(a)$ is straightforward but tedious, and we shall merely give the final result. Equation (6.10) may be written in the form

$$\eta = \eta_0 \left[1 + \frac{5v}{2V} \left\{ 1 + q^* \left(\frac{e^2 Q}{c a k T} \right)^2 Z(b) \right\} \right], \quad (6.14)$$

where
$$q^* = \epsilon k T \sum_{i=1}^s n_i z_i^2 w_i^{-1} / \eta_0 e^2 \sum_{i=1}^s n_i z_i^2, \quad b = \chi a \quad (6.15)$$

and
$$\begin{aligned} Z(b) &= \frac{1}{250\pi(1+b)^2} \left[-\frac{75}{4b^2} - \frac{25}{4b} - \frac{25}{8} + \frac{75b}{8} - \frac{2785b^2}{336} - \frac{265b^3}{168} - \frac{1037b^4}{224} \right. \\ &\quad + \frac{593b^5}{672} - \frac{67b^6}{42} - \frac{b^7}{14} + \frac{b^8}{7} + e^{2b} \text{Ei}(2b) \left(-25b^2 + \frac{25b^5}{3} + \frac{10b^7}{3} - \frac{2b^9}{7} \right) \\ &\quad + e^b \text{Ei}(b) \frac{75}{16} \left(1 - \frac{b^2}{30} \right) (2b^2 + 2b^3 + b^4) + F'_1(b, b) (1+b) \\ &\quad \times \left. \left\{ -\frac{25b}{24} + \frac{5b^2}{8} - \frac{5b^3}{12} + \frac{35b^4}{72} + \frac{5b^5}{288} - \frac{5b^6}{288} - \frac{25}{48} e^b \text{Ei}(b) b^5 \left(1 - \frac{b^2}{30} \right) \right\} \right]. \quad (6.16) \end{aligned}$$

The quantity q^* is dimensionless and is independent of the concentration for an electrolyte of given composition—that is, with given ionic species and given proportions of the various species.

The formula for η in terms of ζ the electrokinetic potential is

$$\eta = \eta_0 \left[1 + \frac{5}{2} \frac{v}{V} \left\{ 1 + q^* \left(\frac{e\zeta}{kT} \right)^2 Z(b) (1+b)^2 \right\} \right]. \quad (6.17)$$

The limiting forms of $Z(b)$ which are useful for b small or b large are respectively

$$Z(b) \doteq \frac{1}{200\pi b} + \frac{11b}{3200\pi} \quad (6.18)$$

and

$$Z(b) \doteq \frac{3}{2\pi b^4}. \quad (6.19)$$

$Z(b)$ is plotted as a function of b ($= \chi a$) in figure 1.

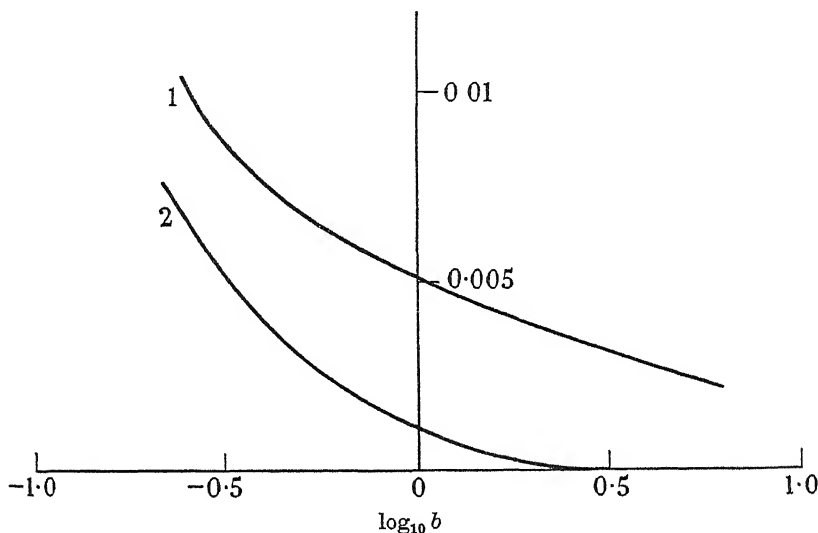


FIGURE 1. Functions $Z(b)$ (curve 2) and $Z(b)(1+b)^2$ (curve 1) plotted against $\log_{10} b$.

7. DISCUSSION

From formula (6.18) it will be noticed that for a given charge or potential the electroviscous contribution to η increases as b^{-1} for small values of b . At first sight this result would seem to be incorrect, since in the limit $b \rightarrow 0$ (small spheres) it would make η tend to infinity. But the limit $b \rightarrow 0$ corresponds also to a double layer extending to infinity, or effectively, a charged sphere in a liquid medium with no free ions; such a system could obviously not behave as a rigid body. However, it is not legitimate to reduce b to zero, since we have assumed in condition (2.7) that the double layer does not extend near to the boundary of the fluid sphere. This assumption was introduced to obviate difficulties due to possible effects of external particles on the double layer of the solid sphere in the sphere $r = r_0$. But it has also been introduced explicitly in the calculation in equation (6.7).

On physical grounds it is to be expected that the electroviscous effect should increase as the thickness of the double layer increases. For, basically, the effect arises from the distortion of the spherically symmetrical field round the particle,

due to the ions being dragged along by the fluid. Since the fluid velocity increases roughly as the distance from the centre as we move away from the sphere, we should expect the distortion to be greater the farther out from the sphere the main body of the ionic charge lies, that is, the more diffuse the double layer.

For b large, or for a very thin double layer, we observe from (6.19) that the electroviscous contribution vanishes. This again is to be expected from general physical arguments. When b is large the double layer is concentrated in a region very near to the surface where the fluid velocity is zero. Hence there should be no distortion of the field and no modification of the velocity components on the surface of the larger sphere $r = r_0$. But this result is in direct contradiction to the Smoluchowski and Krasny-Ergen formulae (1.3), for they apply to a thin double layer and obviously do not vanish. Consequently it is now necessary to undertake a critical examination of Krasny-Ergen's analysis.

In principle, Krasny-Ergen's method is to find the distortion of the potential due to the fluid motion. This potential produces currents in the electrolyte, and it is assumed that the energy dissipated is $\sigma(\text{grad } \Phi)^2$ per unit volume, where σ is the specific conductivity and Φ the distortion of the potential. Hence the total energy dissipated throughout the sphere of radius r_0 of volume τ will be

$$W_{\text{el.}} = \sigma \int_{\tau} (\text{grad } \Phi)^2 d\tau. \quad (7.1)$$

The integral of equation (7.1) is evaluated by relating Φ to p the fluid pressure in Einstein's calculation, it is found that

$$\Phi = \frac{e\xi}{4\pi\eta_0\sigma} p + \text{constant}. \quad (7.2)$$

From the expression for $W_{\text{el.}}$ the effective viscosity is found by Einstein's method; the work done in unit time in the sphere of radius r_0 is equated to the work done in a like sphere of uniform fluid, with viscosity η , and with the same velocity distribution on its boundary $r = r_0$.

There are two criticisms to be made of this argument. In the first place in expression (7.1) the contribution due to the currents caused by diffusion gradients is neglected.

It can also be shown, without a detailed examination of Krasny-Ergen's deduction of equation (7.2), that his relation between Φ and p must be incorrect. For, on the surface of the sphere, from equation (14) of Krasny-Ergen's paper, we have $\partial\Phi/\partial r|_a = 0$; but this boundary condition is not fulfilled by equation (7.2).

8. COMPARISON WITH EXPERIMENT

Unfortunately, very few experimental results seem to be available on the electroviscous effect. However, it is interesting to examine some recent work by Bull (1940) on the viscosity of suspensions of egg albumen. Bull measured the viscosity of the suspension, together with its specific conductivity and the electrophoretic mobility, for various values of pH; in some cases measurements were made with an electrolyte

(sodium chloride) added. He concluded that the electrical contribution to the viscosity varies as ζ^2 as predicted by the Smoluchowski and Krasny-Ergen equations, but that both formulae gave a completely wrong order of magnitude for the effect. The Smoluchowski formula gave values as high as ninety times those observed, and the disparity was obviously greater with the Krasny-Ergen formula.

Let us now compare Bull's data with the calculations of the present paper. Unfortunately, exact comparison is not possible for various reasons. In the first place, the particles of protein which made up the suspension are not spheres but spheroids with a fairly large eccentricity (axial ratio of the order 3). Possibly for this reason Bull found that at the isoelectric point

$$\eta = \eta_0(1 + 5.2v/V), \quad (8.1)$$

instead of the usual Einstein formula. Secondly, formula (6.17) gives η in terms of the ionic mobilities w_i , not the specific conductivity σ . (6.17) can be brought into a suitable form if we assume that all the mobilities are equal. We can then write for the electroviscous contribution to η

$$\eta_{el} = \eta_0 \frac{5.2v}{V} \frac{1}{\sigma \eta_0 a^2} \left(\frac{\zeta \epsilon}{2\pi} \right)^2 \pi b^2 (1+b)^2 Z(b), \quad (8.2)$$

which differs from Smoluchowski's formula by the addition of the multiplying factor $\pi b^2(1+b)^2 Z(b)$. In table 1 we compare Bull's data from formula (8.2).

TABLE 1. CALCULATED AND OBSERVED VALUES OF η_{el}

pH	$\eta_{el} v/V$			
	b	observed	Krasny-Ergen	formula (8.2)
without NaCl				
5.25	0.0186	0.6	60	0.01
5.85	0.00941	1.8	254	0.01
6.62	0.00135	2.9	543	0.02
7.72	0.00594	3.9	934	0.02
9.65	0.0504	4.6	1299	0.2
10.20	0.1661	4.3	1126	0.8
10.85	0.3570	3.2	645	1.4
11.16	0.5101	2.5	424	1.5
with NaCl (0.01N)				
5.71	0.7892	0.4	14.4	0.1
7.38	0.7892	0.6	37.3	0.3
8.82	0.7892	0.5	103.5	0.75
10.60	0.8333	0.3	90.8	0.7

The second column gives the parameter b calculated from the pH values and concentration of added electrolyte. Comparing the third and fifth columns of the table it will be seen that our theory for the most part gives better agreement with experiment than the Krasny-Ergen formula. The agreement is poor in the first six entries. For various reasons, however, too much significance should not be attached to this disparity. For example, Bull states that for the viscosity measurements, concentrations of less than 1 % of protein were used. Hence condition (2.7) puts

a lower limit of 0.1 on the value of b for which the formula (8.2) is valid. It is not, therefore, strictly legitimate to apply our theory to the first five entries of the table. A more likely explanation of the disparity, however, is that the values of b estimated from the pH values are much too small; in other words, other ions were present in the suspension. For the values of b in the first five cases indicate a double layer so diffuse as to make the suspension very nearly equivalent to a suspension of charged spheres in a non-ionizing medium. Such a system would not show any electrical contribution to the viscosity; this is contrary to the experimental facts, and we accordingly conclude that the estimates of b are too small. Yet another error may have been introduced owing to ambiguity in the interpretation of the quantity σ . In formula (8.2), σ refers to the specific conductivity of the *electrolyte only*. Bull's measurements, however, seem to have been made upon the *suspension*; hence part of the specific conductivity he measured may have been due to the protein itself. Any substantial error due to this cause would, of course, improve the agreement of our theory with experiment but increase the disparity between experiment and the Krasny-Ergen theory.

9. APPENDIX

For equations (6.3) to be correct, we must verify that the boundary condition (6.4) at the surface $r = a$ is also correct; to do this we shall show that the total force and resultant couple on the solid sphere both vanish.

The forces on the sphere are of two kinds:

- (a) ordinary stresses over the surface due to the fluid motion and pressure;
- (b) electrical forces due to the interaction of the charges on the surface and the charges in the ionic atmosphere.

Consider the forces of type (a); the appropriate stress components are

$$\left. \begin{aligned} S_{rr} &= -p + 2\eta_0 \frac{\partial v_r}{\partial r}, \\ S_{\theta r} &= \eta_0 \left[\frac{1}{r} \frac{\partial v_r}{\partial \theta} + \frac{\partial v_\theta}{\partial r} - \frac{v_\theta}{r} \right], \\ S_{\alpha r} &= \eta_0 \left[\frac{1}{r \sin \theta} \frac{\partial v_r}{\partial \alpha} + \frac{\partial v_\alpha}{\partial r} - \frac{v_\alpha}{r} \right]. \end{aligned} \right\} \quad (9.1)$$

$S_{\theta r}$ is the stress over unit area (r constant), in the direction of θ increasing; the other stresses are defined similarly.

Resolving along the (x, y, z) axes we have

$$\left. \begin{aligned} S_{xr} &= S_{rr} \sin \theta \cos \alpha + S_{\theta r} \cos \theta \cos \alpha - S_{\alpha r} \sin \alpha, \\ S_{yr} &= S_{rr} \sin \theta \sin \alpha + S_{\theta r} \cos \theta \sin \alpha + S_{\alpha r} \cos \alpha, \\ S_{zr} &= S_{rr} \cos \theta - S_{\theta r} \sin \theta. \end{aligned} \right\} \quad (9.2)$$

If S_x denotes the total force on the solid sphere in the x direction due to the stresses then

$$S_x = a^2 \int_0^\pi \int_0^{2\pi} (S_{xr})_{r=a} \sin \theta d\theta d\alpha. \quad (9.3)$$

From (9.1), (9.2), (5.43) and (5.46) this becomes

$$\begin{aligned}
 S_x = \eta_0 a^2 \left[-\frac{3A}{7} a^2 + \frac{6B}{a^2} + 4C + \frac{12D}{a^3} + 2X''(a) - U(a) - aU'(a) \right. \\
 \left. - \frac{\epsilon}{20\pi\eta} \int_a^\infty \left(2\Delta_0^{(r)} \psi_1(\chi^r) \Delta_2^{(r)} \Psi(r) + \frac{d\Psi(r)}{dr} \frac{d}{dr} \Delta_0^{(r)} \psi_1(\chi^r) + \frac{d\psi_1(\chi^r)}{dr} \frac{d}{dr} \Delta_2^{(r)} \Psi(r) \right) \right. \\
 \left. \times \left(\frac{r^4}{a^3} - \frac{a^2}{r} \right) dr \right] \int_0^\pi \int_0^{2\pi} f \sin^2 \theta \cos \alpha d\theta d\alpha \\
 + \eta_0 a^2 \left[\frac{8A}{7} a^2 + 2C - \frac{8D}{a^5} + \frac{2}{a} X'(a) - U(a) - \frac{2}{a^2} X(a) \right] \\
 \times \int_0^\pi \int_0^{2\pi} \left(\frac{\partial f}{\partial \theta} \sin \theta \cos \theta \cos \alpha - \frac{\partial f}{\partial \alpha} \sin \alpha \right) d\theta d\alpha. \quad (9.4)
 \end{aligned}$$

Both integrals on the right-hand side of this equation are zero. Hence it follows that

$$S_x = 0. \quad (9.5)$$

Similarly, it can be shown that $S_y = S_z = 0$.

If L_x , L_y and L_z are the couples due to the surface stresses, we find

$$\begin{aligned}
 L_x = a^2 \int_0^\pi \int_0^{2\pi} (zS_{ry} - yS_{rz}) \sin \theta d\theta d\alpha \\
 = \eta_0 a^2 \left[\frac{8A}{7} a^2 + 2C - \frac{8D}{a^5} + \frac{2X'(a)}{a} - U(a) + \frac{2X(a)}{a^2} \right] \\
 \times \int_0^\pi \int_0^{2\pi} \left(\frac{\partial f}{\partial \theta} \sin \theta \sin \alpha + \frac{\partial f}{\partial \alpha} \cos \alpha \cos \theta \right) d\theta d\alpha, \quad (9.6)
 \end{aligned}$$

since the integral over the angle variables again vanishes. Similarly we can show that $L_y = L_z = 0$.

Now consider the forces of type (b). It is convenient to regard the total potential ψ as the sum of two potentials ψ_e and ψ_s , ψ_e due to the ions in the electrolyte, the other due to the surface charge. For the surface charge density σ_s we have

$$\sigma_s = -\frac{1}{4\pi} \epsilon \frac{d\psi_1}{dr} \Big|_a. \quad (9.7)$$

ψ_s in the electrolyte satisfies Laplace's equation, and it is easily shown that

$$\psi_s = \psi_1 + \frac{a^2}{r} \frac{d\psi_1}{dr} \Big|_a + f \left(\Psi + \frac{a^4}{3r^3} \frac{d\Psi}{dr} \Big|_a \right). \quad (9.8)$$

From (9.7) and (9.8) the total forces and moments acting on the particle due to the electrostatic field are easily determined. For example, the force in the x direction is

$$E_x = -a^2 \int_0^\pi \int_0^{2\pi} \sigma_s \frac{\partial \psi_e}{\partial x} \Big|_a \sin \theta d\theta d\alpha. \quad (9.9)$$

Using (9.7) and (9.8) it is easily verified that the integrals over the angle variables vanish.

Hence

$$E_x = 0.$$

Similarly

$$E_x = E_y = 0.$$

The couples on the sphere can also be shown to vanish. We therefore conclude that there is no resultant thrust or moment acting on the sphere, and therefore that boundary condition (6.4) is valid.

I am indebted to Imperial Chemical Industries Ltd. for a research grant under which this work was carried out.

REFERENCES

- Booth, F. 1950 *Proc. Roy. Soc. A*, **203**, 514.
 Bull, H. B. 1940 *Trans. Faraday Soc.* **36**, 80.
 Einstein, A. 1906 *Ann. Phys., Lpz.*, **19**, 289.
 Einstein, A. 1911 *Ann. Phys., Lpz.*, **34**, 591.
 Frohlich, H. & Sack, R. 1946 *Proc. Roy. Soc. A*, **185**, 415.
 Guth, E. & Simha, R. 1936 *Kolloidzshr* **74**, 266.
 Jeffery, G. B. 1922 *Proc. Roy. Soc. A*, **102**, 161.
 Krasny-Ergen, W. 1936 *Kolloidzshr* **74**, 172.
 Smoluchowski, M. 1916 *Kolloidzshr* **18**, 190.

Exact solutions for flow of a perfect gas in a two-dimensional Laval nozzle

By T. M. CHERRY

(Communicated by S. Goldstein, F.R.S.—Received 27 May 1949—
 Revised 10 March 1950)

A family of exact solutions is found for the problem of steady irrotational isentropic shock-free transsonic flow of a perfect gas through a Laval nozzle in two dimensions. The hodograph method is used, whereby the position co-ordinates x, y are expressed in terms of the velocity variables; the expressions are infinite series in the subsonic part of the flow field, infinite integrals (analytic continuations of the series) in the supersonic part. An inversion is required to get the velocity as a function of position, in general, this requires detailed numerical calculations, but approximate formulae (62) are found for the neighbourhood of the sonic point on the axis.

INTRODUCTION

The hodograph method, which has been described by many authors, for example, Cherry (1947), is used, taking the usual independent variables τ , proportional to the square of the speed, and θ , the inclination of the velocity vector to the x -axis. The Legendre potential Ω satisfies a linear differential equation having solutions of the form

$$\Omega = \sum_{\nu} A_{\nu} \chi_{\nu}(\tau) e^{\pm i\nu\theta}, \quad (1)$$

where $\chi_\nu(\tau)$ is a hypergeometric function defined in appendix 1; and from Ω the position co-ordinates x, y are obtained by differentiation

$$x\tau^{\frac{1}{2}}/a = 2\tau\Omega_\tau \cos \theta - \Omega_\theta \sin \theta, \quad y\tau^{\frac{1}{2}}/a = 2\tau\Omega_\tau \sin \theta + \Omega_\theta \cos \theta, \quad (2)$$

where a is an arbitrary scale constant.

It is well known that most specific problems of gas flow, i.e. ones that demand a unique solution, are intractable by the hodograph (or any other) method. The most that we can usually achieve is the construction of flows that are, somewhat vaguely, of a 'prescribed type', e.g. 'flow past a cylinder'; and what we shall here find are flows of the 'Laval type', defined as follows:

(i) *The flow is steady, irrotational, isentropic and two-dimensional.*

(ii) *There is a straight streamline Ox , which is an axis of symmetry for the flow field. On this axis the speed increases steadily with x , being subsonic for $x < x_s$, supersonic for $x > x_s$.*

(iii) *The velocity co-ordinates τ, θ are analytic functions of x, y throughout the field; in particular, no shocks are present.*

It is known (Lighthill 1947*a*) that, when conditions (i) to (iii) are satisfied, there is a locus through the axial sonic point O (d, d' in figure 1) on which

$$J = \frac{\partial(\tau, \theta)}{\partial(x, y)} = 0.* \quad (3)$$

By (iii) J is an analytic function of x, y , and our last prescribed condition is

(iv) *On the locus (d, d'), $J = 0$, but $\partial(\tau, J)/\partial(x, y) \neq 0$; this condition is simply that we have the 'most general' case.*

The essential problem is to find one such flow; once this has been done, it is easy to specify a family of them, depending on an unlimited number of adjustable parameters.

1. GENERAL SKETCH OF THE INVESTIGATION

The main steps are:

I. From conditions (i) to (iv) above, we deduce the character which x, y , and thence Ω , must have as functions of τ, θ . The leading feature is that, for values of τ corresponding to supersonic speeds, Ω is three-valued, with two real branch lines (§2).

II. We guess a series of the form (1), which seems likely to define a function Ω having the necessary character (§§3, 7).

III. We verify that the chosen Ω has this character (§§4 to 7).

IV. Inserting the chosen Ω in (2), we verify that these equations can be solved for τ, θ in terms of x, y and that the conditions (i) to (iv) for flow of the Laval type are satisfied (§§8, 9).

The deductive part of the investigation lies entirely in III, IV, the steps I, II are analysis, telling us where to start the deduction.

* The existence of such a locus is almost evident from the fact that $\theta = 0$ not merely on the axis Ox (figure 1) but also on a locus c, c' passing through O .

In general plan, therefore, the investigation is similar to the solutions which Lighthill (1947*b*) and myself (Cherry 1947) have given for the problem of gas flow past a cylinder; but the present problem is more difficult. In the cylinder problem, the correct guess (step II) was almost obvious from a consideration of the limiting case in which the fluid becomes incompressible; but the Laval problem has, from its nature, no such limiting case. The guess was guided by the observation that the hypergeometric function $\chi_\nu(\tau)$ in (1) has very similar properties to the Bessel function

$$(\nu/e)^{-\nu} \Gamma(\nu+1) \delta^\nu J_\nu(\nu t), \quad (4)$$

provided δ is a suitably chosen constant and t a suitably chosen function of τ . Hence we may expect the series (1) to have similar properties to Kapteyn series

$$\Sigma B_\nu J_\nu(\nu t) e^{\pm \nu \theta}.$$

Now Kapteyn series have been extensively studied. Some of them can be summed in closed form, and then it is easy to determine the singularities of the sum function. It was found that the very simplest such series

$$\sum_{-\infty}^{\infty} J_n(nt) e^{in\theta},$$

defines a function having almost precisely the character desired in Ω , and thence the likely guess for Ω (which was in fact successful) was evident.

Step III—the verification that the chosen Ω has the desired character—is the most intricate part of the investigation. The difficulty is that a series of the form (1) can represent Ω only in ranges of τ, θ where it is non-singular, so we require a method for obtaining analytic continuations of such series, valid in new ranges of τ, θ . Here again the method—transformation of (1) into a new series whose individual terms submit fairly readily to analytic continuation—is analogous to the one used in the cylinder problem, but not so easy. In the cylinder problem, each individual term of the transformed series was an ‘elementary’ series; but in the present problem it is a Kapteyn series.

2. THE SINGULARITIES OF THE POTENTIAL Ω

The implications of the conditions (i) to (iv), defining a flow of the Laval type, have been investigated by Lighthill (1947*a*). The part of the xy -plane (physical plane) near O , the sonic point on the axis, is in one-one correspondence with a three-sheeted Riemann surface for τ, θ (hodograph surface), as indicated in figures 1, 2. Corresponding domains and curves are here similarly named. $CDEF$ and $C'D'E'F'$ are non-axial streamlines; the curves a, a' are the sonic locus $\tau = \tau_s$; c, c' are the non-axial part of the locus $\theta = 0$; b, d' are characteristics $\theta = \omega(\tau)$; and b', d are characteristics $\theta = -\omega(\tau)$, where

$$\omega(\tau) = \tau_s^{-\frac{1}{2}} \arctan \sqrt{\frac{\tau - \tau_s}{1 - \tau}} - \arctan \sqrt{\frac{\tau/\tau_s - 1}{1 - \tau}}$$

(τ_s gives the sonic speed, and is defined in (7) below.) The Jacobian J , equation (3), vanishes on d and d' , but not on b, b' .

Let (x_0, y_0) be a point on d or d' , with corresponding values τ_0, θ_0 . By condition (iii), $\tau - \tau_0$, $\theta - \theta_0$ and J are power series in $x - x_0$, $y - y_0$ vanishing with these arguments, and by condition (iv) we can solve for $x - x_0$, $y - y_0$ in powers of $\tau - \tau_0$ and J :

$$x - x_0 = P_1(\tau - \tau_0, J), \quad y - y_0 = P_2(\tau - \tau_0, J). \quad (5)$$

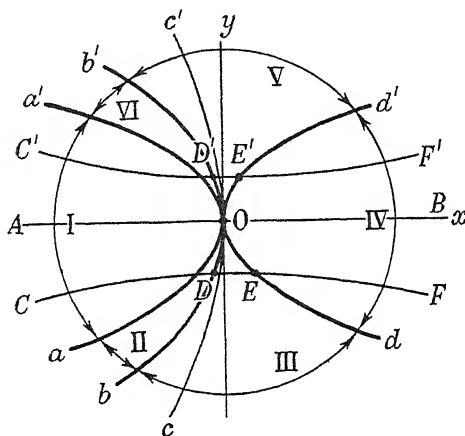


FIGURE 1. Physical plane.

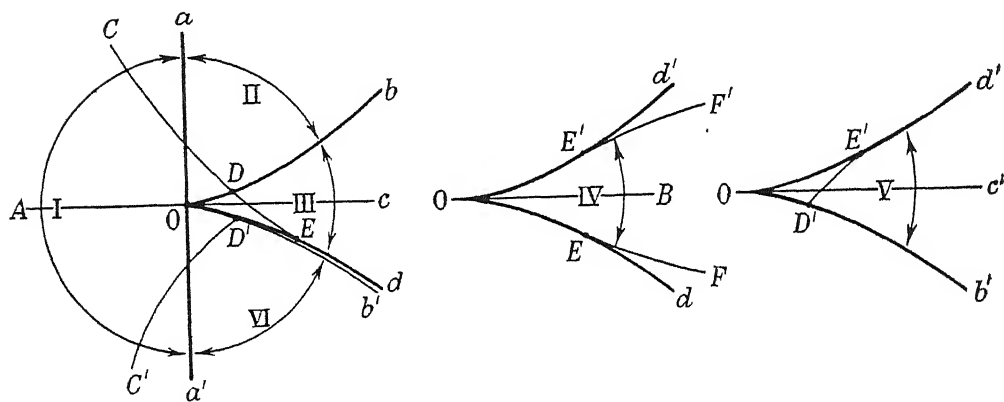


FIGURE 2. Hodograph surface (τ -axis horizontal, θ -axis vertical).

Thence $\theta - \theta_0$ is a similar series:

$$\theta - \theta_0 = P(\tau - \tau_0, J); \quad (6)$$

and here

$$P_J = \frac{\partial(\tau, \theta)}{\partial(x, y)} \bigg/ \frac{\partial(\tau, J)}{\partial(x, y)} = J \bigg/ \frac{\partial(\tau, J)}{\partial(x, y)},$$

where by condition (iv) the denominator is non-zero for $\tau = \tau_0$, $J = 0$. So (6) must have the form

$$\theta - \theta_0 = P(\tau - \tau_0, 0) + J^2 Q(\tau - \tau_0, J),$$

with $Q(0, 0) \neq 0$; we can solve for J in powers of $\tau - \tau_0$ and $\sqrt{\{\theta - \theta_0 - P(\tau - \tau_0, 0)\}}$, and the radical must be $\sqrt{(\theta + \omega)}$ on characteristic d , $\sqrt{(\theta - \omega)}$ on characteristic d' .

Substituting for J in (5), x, y become power series in $\tau - \tau_0$ and $\sqrt{(\theta \pm \omega)}$; and by (2) Ω must be a similar series, except that the lowest odd power of the radical must be $(\theta \pm \omega)^{\frac{3}{2}}$.

The characterization of Ω is therefore as follows:

(A) For a fixed supersonic τ let Ω be regarded as a function of the complex variable θ . Then at any point of the locus d , where $\theta = -\omega(\tau)$, Ω has a simple branch point; both branches are real when $\theta + \omega(\tau)$ is real positive, and one of them specifies Ω in region III, the other in region IV. There are similar statements for the neighbourhood of the branch line d' , where $\theta = \omega(\tau)$.

(B) At any point not on the branch lines d , d' , Ω is a regular function of τ, θ ; in particular, this is so on the characteristics b , b' , where again $\theta = \pm \omega(\tau)$.

(C) The equations (2) are to be soluble for τ, θ as regular functions of x, y satisfying conditions (i) to (iv) of §1.

3. THE BESSEL FUNCTION ANALOGY, FUNDAMENTAL FORMULAE

It has been explained in §1 that the key for all that is to follow lies in an analogy between the hypergeometric function $\chi_\nu(\tau)$ and the Bessel function (4). There is, indeed, more than an analogy, the one function may be expressed in terms of the other, as follows:

THEOREM.* Let $\beta, \alpha, \delta, \tau_s$ be constants depending on the adiabatic index γ of the gas, as follows:

$$\beta = \frac{1}{\gamma-1}, \quad 2\alpha = \sqrt{(1+2\beta)-1}, \quad \delta = \frac{\alpha^\alpha}{(1+\alpha)^{1+\alpha}}, \quad \tau_s = \frac{1}{1+2\beta} = \frac{\gamma-1}{\gamma+1}. \quad (7)$$

Let t be determined as a function of τ by

$$\sqrt{(t^2-1)} - \arctan \sqrt{(t^2-1)} = \omega(\tau) = \tau_s^{-\frac{1}{2}} \arctan \sqrt{\frac{\tau-\tau_s}{1-\tau}} - \arctan \sqrt{\frac{\tau/\tau_s-1}{1-\tau}}, \quad (8)$$

and let

$$Q = Q(\tau) = (1-\tau)^{\frac{1}{2}-\frac{1}{2}\beta} \left(\frac{t^2-1}{\tau/\tau_s-1} \right)^{\frac{1}{2}}. \quad (9)$$

Then for $0 \leq \tau < 1$ and all complex ν with $-\pi \leq \arg \nu \leq \pi$,

$$\frac{(\nu/e)^\nu \chi_\nu(\tau)}{\delta^\nu \Gamma(\nu+1)} = J_\nu(\nu t) \left\{ Q(\tau) + \sum_\mu \frac{f_\mu(\tau)}{\nu-\mu} \right\}, \quad (10)$$

where the summation is over the zeroes $\mu = \mu(\tau)$ of $(\mu/e)^{-\mu} J_\mu(\mu t)$, and $f_\mu(\tau)$ depends on μ, τ but not on ν .

Note 1. If we multiply both sides of (10) by $(\nu/e)^{-\nu} \tau^{-\frac{1}{2}\nu}$, they become single-valued throughout the ν -plane and throughout the τ -plane cut from $\tau = 1$ to $+\infty$; and in this form the theorem is true for all complex ν, τ . For present purposes, however, the form (10) is what is directly required.

Note 2. In (8) it is understood that principal values of the roots and inverse tangents are taken for $\tau > \tau_s$ and $t > 1$; then as τ increases from τ_s to 1, t increases from 1 to a finite value t_1 . For any complex τ which is not on the cuts $(1, +\infty)$ or $(0, -\infty)$ t is thence determinate by analytic continuation, there being no singularity at $\tau = \tau_s$. For τ near τ_s there is a regular development

$$t-1 = \frac{\tau-\tau_s}{2\tau_s(1-\tau_s)^{\frac{1}{2}}} + O(\tau-\tau_s)^2. \quad (11)$$

* For proof, see appendix 2.

For τ near 0, $t/\sqrt{\tau}$ is regular: $t/\sqrt{\tau} = 2/(\epsilon\delta) + O(\tau)$.

Figure 3 shows the graph of (8) in the case where $\gamma = 1.4$.

By (9) and (11) Q is regular near $\tau = \tau_s$, $t = 1$, its value at this point being

$$Q_s = (1 - \tau_s)^{\frac{1}{2} - \frac{1}{2}\beta}. \quad (12)$$

Note 3. For $0 < \tau < \tau_s$ all the zeroes μ are real negative, the large ones being a little greater than the negative integers $-n$. For $\tau = \tau_s$ this remains true, except that the large zeros are now $-n + \frac{1}{6}$, approximately. For $\tau_s < \tau < 1$ the zeros μ are all real; there is a positive set

$$\mu = \mu_n = (n - \tfrac{1}{4})\pi/\omega + O(1/n), \quad \text{giving } f_{\mu_n}(\tau) = O(1/n), \quad (13a)$$

and a negative set

$$\mu = \mu'_n = -(n - \tfrac{1}{4})\pi/(\pi + \omega) + O(1/n), \quad \text{giving } f_{\mu'_n}(\tau) = -f_{\mu_n}(\tau) + O(n^{-2}). \quad (13b)$$

(The order of magnitude of $f_{\mu}(\tau)$ in the case $0 < \tau \leq \tau_s$ is given in appendix 2; we shall not need to use the theorem in this case.)

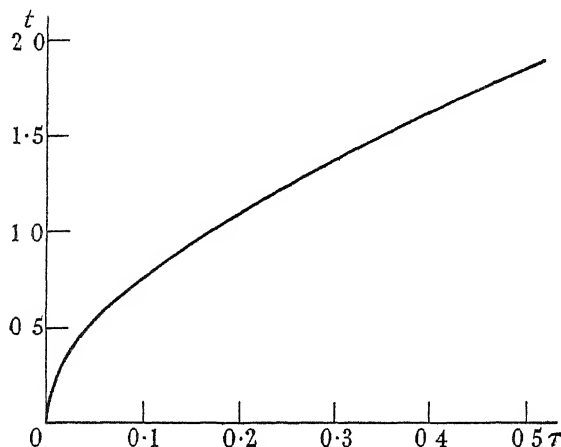


FIGURE 3. The t, τ relation.

If in (10) we replace ν by $\nu e^{\pi i}$ and use $J_{-\nu}(ze^{\pi i}) = e^{-\nu\pi i} J_{-\nu}(z)$, we obtain

$$\frac{\Gamma(\nu) \sin \nu\pi \chi_{-\nu}(\tau)}{\pi \delta^{-\nu} (\nu/e)^{\nu}} = J_{-\nu}(\nu t) \left\{ Q - \sum_{\mu} \frac{f_{\mu}}{\nu + \mu} \right\}. \quad (14)$$

Let $\nu \rightarrow n$, a positive integer. Here $\chi_{-\nu}(\tau)$ has a pole at which its residue is $h_n \chi_n(\tau)$, where h_n is a positive constant defined in appendix 1, so we obtain

$$\frac{h_n \delta^n \Gamma(n) \chi_n(\tau)}{(n/e)^n} = J_n(nt) \left\{ Q - \sum_{\mu} \frac{f_{\mu}}{n + \mu} \right\}. \quad (15)$$

From (10) and (15) there follows

$$\begin{aligned} \sum_0^{\infty} \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{\delta^n \Gamma(n+1)} + \sum_1^{\infty} \frac{h_n \delta^n \Gamma(n) \chi_n(\tau) e^{in\theta}}{(n/e)^n} &= Q \left\{ 1 + \sum_1^{\infty} 2J_n(nt) \cos n\theta \right\} \\ &\quad - \sum_{\mu} f_{\mu} \left\{ \frac{1}{\mu} + \sum_{n=1}^{\infty} J_n(nt) \left(\frac{e^{-in\theta}}{\mu - n} + \frac{e^{in\theta}}{\mu + n} \right) \right\}, \end{aligned} \quad (16)$$

$$= QK(t, \theta) - \sum f_{\mu} L_{\mu}(t, \theta), \quad \text{say}, \quad (16a)$$

provided the rearrangement of the double series on the right is legitimate.

In (16), we have chosen the series on the left so as to secure for the coefficient of Q , on the right, a well-known Kapteyn series; when $0 < t < 1$ and θ is real its sum is

$$K(t, \theta) = 1 + \sum_{n=1}^{\infty} 2J_n(nt) \cos n\theta = 1/(1 - t \cos \phi), \quad (17)$$

where ϕ is the unique real root of

$$\phi - t \sin \phi = \theta. \quad (18)$$

If, now, t, θ vary continuously in any manner, the root ϕ of (18) varies continuously in a manner which is in general determinate, and the expression on the right of (17) gives the analytic continuation of the function $K(t, \theta)$; the determinateness breaks down only when ϕ becomes a multiple root of (18), and when t, θ are such that this occurs $K(t, \theta)$ has a branch point. This branch condition is $t = \sec \phi$, and hence

$$\theta = \pm \{\arctan \sqrt{(t^2 - 1)} - \sqrt{(t^2 - 1)}\} = \mp \omega(\tau), \quad \text{by (8)}. \quad (19)$$

A complete view of what happens for all real t, θ is obtained when we note that (18) sets up one-one correspondence between the $t\phi$ -plane and a three-sheeted Riemann surface for t, θ , for which (19) gives the branch lines; so this surface is precisely that shown in figure 2. The $t\phi$ -plane, with the loci and domains corresponding to those of figure 2 or figure 1, is shown in figure 4.

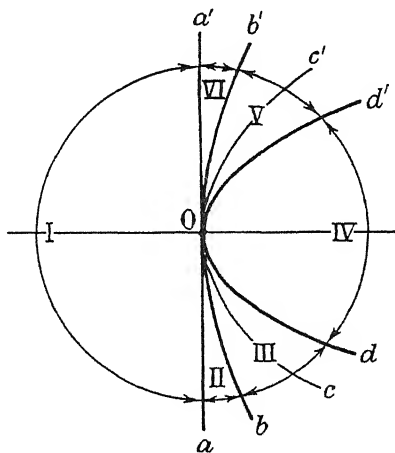


FIGURE 4. The $t\phi$ -plane (t -axis horizontal, ϕ -axis vertical).

Hence, since $K(t, \theta)$ is a single-valued function of t, ϕ , it has, as regards t, θ , just the sort of triple-valuedness which we want in Ω , according to conditions (A), (B) at the end of §2. The only discrepancy is that K is infinite on the branch lines, whereas we require Ω to be finite, but integration with respect to θ will cure that (see §7).

The suggestion therefore is that the left-hand side of (16) may (after integration) give the solution Ω which we are seeking. We shall prove this by finding the appropriate analytic continuations of the Kapteyn series on the right of (16). We shall work with the variables t, θ (not t, ϕ),* using the standard method of deriving a chain of expressions such that (i) each represents an analytic function in some domain, (ii) the domains associated with two successive expressions of the chain overlap,

* The ultimate justification for abandoning ϕ lies in the compact final formulae of §§6, 7.

(iii) these two expressions are equal throughout the region of overlap. Though our ultimate interest is only in real t, θ , the theory requires that they be allowed to vary, independently, in strips containing the real axes in their respective complex planes. The 'domains' just spoken of are then four-dimensional, but for the truth of (iii) it is sufficient that the expressions be equal on the real two-dimensional region, contained in the four-dimensional overlap, given by $t_1 < t < t_2, \theta_1 < \theta < \theta_2$.

The following properties of Bessel functions will be required:

$$(i) J_\nu(z) = \frac{1}{2}\{H_\nu^{(1)}(z) + H_\nu^{(2)}(z)\}, \quad J_{-\nu}(z) = \frac{1}{2}\{e^{\nu\pi i}H_\nu^{(1)}(z) + e^{-\nu\pi i}H_\nu^{(2)}(z)\}. \quad (20)$$

$$(ii) J_\nu(ze^{\pm\pi i}) = e^{\pm\nu\pi i}J_\nu(z), \quad H_{-\nu}^{(1)}(ze^{\pi i}) = -H_\nu^{(2)}(z), \quad H_{-\nu}^{(2)}(ze^{-\pi i}) = -H_\nu^{(1)}(z). \quad (21)$$

(iii) For $\nu \sim 0$, uniformly for $\text{Rl } t \geq 0$ with $|t|$ bounded from 0 and ∞ ,

$$J_\nu(\nu t) = 1 + O(\nu \log \nu), \quad H_\nu^{(r)}(\nu t) = O(\log \nu) \quad (r = 1, 2). \quad (22)$$

(iv) *Behaviour for $\nu \sim \infty$.* The following inequalities, valid actually for all $|\nu|$, will be sufficient; as regards the regions D_1, D_2 of figure 5, all that matters is that they contain the indicated segments of the real t -axis.

(a) For $|\arg \nu| \leq \frac{1}{2}\pi$ and t in the region D_1 of figure 5,

$$|J_\nu(\nu t)| \leq A |\nu^{-\frac{1}{2}} e^{-\nu u}|, \quad (23)$$

where $u = \text{arc tanh } \sqrt{(1-t^2)} - \sqrt{(1-t^2)} = i\omega$, $\text{Rl } u > 0$, and A is an absolute constant.

(b) For $-\frac{1}{2}\pi \leq \arg \nu \leq \pi$ and t in the region D_2 of figure 5,

$$|H_\nu^{(1)}(\nu t)| \leq A |\nu^{-\frac{1}{2}} e^{\nu\omega}|, \quad (24)$$

where ω is defined in terms of t by (8), and $\text{Rl } \omega > 0$.

(c) For $-\pi \leq \arg \nu \leq \frac{1}{2}\pi$ and t in D_2 ,

$$|H_\nu^{(2)}(\nu t)| \leq A |\nu^{-\frac{1}{2}} e^{-i\nu\omega}|. \quad (25)$$

(d) For real $t \geq 1$, (24) and (25) are valid for $|\arg \nu| \leq \pi$.

(e) For $|\arg \nu| \leq \frac{1}{2}\pi$,

$$|J_\nu(\nu t)| \leq A |\nu^{-\frac{1}{2}}| e^{|\nu u|} \quad (u \rightarrow 0 \text{ as } t \rightarrow 1); \quad (26)$$

this provides for values of t near 1 which are not in D_1 or D_2 .

4. ANALYTIC CONTINUATION OF $K(t, \theta)$

(i) We start with the function defined, for real θ and $0 < t < 1$,* by

$$K = K(t, \theta) = 1 + 2 \sum_{n=1}^{\infty} J_n(nt) \cos n\theta = K_I,$$

the symbol K will be used for any continuation of the function, and K_I, K_{II} , etc., for continuations valid in regions I, II, etc. By (23) the series converges uniformly when t is in D_1 (figure 5), but not in a strip bordering its upper and lower boundaries, and at the same time $|\text{Im } \theta|$ is less than some positive bound; so $K(t, \theta)$ is an analytic function of the two variables.

* The statement of this inequality implies that t is to be real; and similarly in other cases below.

(ii) Now for $0 \leq t \leq 1$ and $0 < \theta < 2\pi$ we have, by (23),

$$\begin{aligned} \sum_1^\infty J_n(nt) e^{-in\theta} &= \frac{1}{2i} \int_\infty^{(1+)} \frac{e^{iv(\pi-\theta)}}{\sin v\pi} J_\nu(vt) dv \\ &= \frac{1}{2i} \int_{\infty i}^{-\infty i} \frac{e^{iv(\pi-\theta)}}{\sin v\pi} J_\nu(vt) dv, \end{aligned} \quad (27)$$

where the path crosses the real axis between $\nu = 0$ and $\nu = 1$. And for any t near 1, by (26), the last integral converges uniformly provided $|u| + \epsilon \leq \text{Re } \theta \leq 2\pi - |u| - \epsilon$. Hence this integral gives a continuation of the series on the left, for t in any neighbourhood $|t-1| < r$ and θ in a strip $\eta(r) < \text{Re } \theta < 2\pi - \eta(r)$, where $\eta(r) \rightarrow 0$ as $r \rightarrow 0$.

Also, appealing to (24), (25), we can for $t \geq 1$ and $\omega(t) < \theta < 2\pi - \omega(t)^*$ reconvert the integral on the right of (27) into the series on the left, so the series continues to represent the same function as t increases through 1, provided θ is restricted as stated.

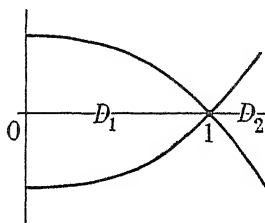


FIGURE 5. t -plane.

Moreover, for $t \geq 1$ and $\omega < \theta < 2\pi - \omega$, we can, by (iv) (d) at the end of §3, rotate the two ends of the path of integration until they lie along the rays $\arg \nu = \pm \pi$, avoiding the points $\nu = 0, -1, -2, \dots$, by indentations. Using residues to find the contributions from the indentations ($((22)_1$ being relevant in the case of $\nu = 0$), we obtain

$$\begin{aligned} \sum_1^\infty J_n(nt) e^{-in\theta} &= \frac{1}{2i} P \int_\infty^0 \frac{J_{-\lambda}(\lambda t e^{\pi i}) e^{-i\lambda(\pi-\theta)} d\lambda}{\sin \lambda\pi} - 1 - \frac{1}{2} \sum_1^\infty J_n(nt) e^{in\theta} \\ &\quad + \frac{1}{2i} P \int_0^\infty \frac{J_{-\lambda}(\lambda t e^{-\pi i}) e^{-i\lambda(\pi-\theta)} d\lambda}{\sin \lambda\pi} - \frac{1}{2} \sum_1^\infty J_n(nt) e^{in\theta} \\ &= \int_0^\infty J_{-\nu}(vt) e^{-iv(\pi-\theta)} d\nu - 1 - \sum_1^\infty J_n(nt) e^{in\theta}, \end{aligned} \quad (28)$$

where at the last step (21)₁ has been used.

There are similar results regarding the analytic continuation of $\sum_1^\infty J_n(nt) e^{in\theta}$, and combining the two: the function K can be continued through

$$|t-1| < r, \quad \eta(r) < \text{Re } \theta < 2\pi - \eta(r),$$

and for $t \geq 1$ and $\omega < \theta < 2\pi - \omega$, its continuation is

$$K_{\text{II}} = 1 + 2 \sum_1^\infty J_n(nt) \cos n\theta = \int_0^\infty J_{-\nu}(vt) e^{iv(\pi-\theta)} d\nu = \int_0^\infty J_{-\nu}(vt) e^{-iv(\pi-\theta)} d\nu. \quad (29)$$

* The two inequalities for θ are consistent when t is not too much greater than 1, and, indeed, we can prove from (8) that they are consistent for all τ in $\tau_s < \tau < 1$, since then $0 < \omega < \pi$.

(iii) We consider now the continuation of the integral on the right of (29). It will simplify the exposition to suppose t fixed at a real value greater than 1; the modifications needed to cover the case where t can range over D_2 (figure 5) are essentially trivial.

Splitting $J_{-\nu}$ in accordance with $(20)_2$ we have

$$\int_0^\infty J_{-\nu}(\nu t) e^{-i\nu(\pi-\theta)} d\nu = \frac{1}{2} \int_0^\infty H_\nu^{(1)}(\nu t) e^{i\nu\theta} d\nu + \frac{1}{2} \int_0^\infty H_\nu^{(2)}(\nu t) e^{-2i\nu\pi+i\nu\theta} d\nu; \quad (30)$$

by (22) both the integrals on the right converge at $\nu = 0$, uniformly as to θ . By (24), $H_\nu^{(1)}(\nu t) e^{i\nu\theta} = O(\nu^{-\frac{1}{2}} e^{i\nu(\theta+\omega)})$, so if $\theta + \omega$ is real positive we have

$$\int_0^\infty H_\nu^{(1)}(\nu t) e^{i\nu\theta} d\nu = \int_0^\infty e^{i\alpha} H_\nu^{(1)}(\nu t) e^{i\nu\theta} d\nu,$$

for any α between $0, \pi$ inclusive. Here the integral on the right represents an analytic function of θ in the half-plane $-\alpha < \arg(\theta + \omega) < \pi - \alpha$; and as α increases from 0 to π this half-plane rotates clockwise about the pivot $\theta = -\omega$, from an initial position which is the upper half-plane $\text{Im } \theta > 0$ to a final position which is the lower half-plane $\text{Im } \theta < 0$.

$$\text{Similarly,} \quad \int_0^\infty H_\nu^{(2)}(\nu t) e^{-2i\nu\pi+i\nu\theta} d\nu = \int_0^\infty e^{-i\alpha} H_\nu^{(2)}(\nu t) e^{-2i\nu\pi+i\nu\theta} d\nu,$$

where we can take $0 \leq \alpha \leq \pi$ provided $2\pi + \omega - \theta$ is real and positive; and the right-hand integral is (as a function of θ) analytic in a half-plane which swings anti-clockwise, about the point $\theta = 2\pi + \omega$ as pivot, from the upper half-plane as initial position to the lower half-plane as final position.

$$\text{Hence} \quad \int_0^\infty J_{-\nu}(\nu t) e^{-i\nu(\pi-\theta)} d\nu, = K_{\text{III}} \text{ say,} \quad (31)$$

which ostensibly defines a function of θ regular in the upper half-plane, actually gives a function which can be continued across the segment $-\omega < \theta < 2\pi + \omega$ into the lower half-plane; and here it is represented by

$$\begin{aligned} & \frac{1}{2} \int_0^\infty e^{i\pi t} H_\nu^{(1)}(\nu t) e^{i\nu\theta} d\nu + \frac{1}{2} \int_0^\infty e^{-i\pi t} H_\nu^{(2)}(\nu t) e^{-2i\nu\pi+i\nu\theta} d\nu \\ &= -\frac{1}{2} \int_0^\infty H_{-\lambda}^{(1)}(\lambda t e^{\pi i}) e^{-i\lambda\theta} d\lambda - \frac{1}{2} \int_0^\infty H_{-\lambda}^{(2)}(\lambda t e^{-\pi i}) e^{2i\lambda\pi-i\lambda\theta} d\lambda \\ &= \int_0^\infty J_{-\nu}(\nu t) e^{i\nu(\pi-\theta)} d\nu = K'_{\text{III}}. \end{aligned} \quad (32)$$

At the last step we have used $(21)_{2,3}$ and $(20)_2$; and the notation $K_{\text{III}}, K'_{\text{III}}$ is justified on account of the regularity in $-\omega < \theta \leq \omega$. Incidentally, we have verified the last equality in (29).

(iv) Consider now the continuation of K_{III} from the upper half-plane across the ray $\theta < -\omega$. We obtain it by rotating the paths, for both the integrals on the right of (30), into the position $\arg \nu = -\pi$. This gives

$$\begin{aligned} K &= \int_0^\infty e^{-i\pi t} J_{-\nu}(\nu t) e^{-i\nu(\pi-\theta)} d\nu = - \int_0^\infty J_\lambda(\lambda t e^{-\pi i}) e^{i\lambda(\pi-\theta)} d\lambda \\ &= - \int_0^\infty J_\nu(\nu t) e^{-i\nu\theta} d\nu = K_{\text{IV}}, \end{aligned} \quad (33)$$

by (21)_I. The expressions K_{IV} , K'_{III} will be shown in §7 to be unequal; we must therefore describe K_{IV} as the continuation of K into a half-plane $\pi < \arg(\theta + \omega) < 2\pi$ distinct (in the Riemann-surface sense) from the half-plane $-\pi < \arg(\theta + \omega) < 0$ in which $K = K'_{III}$.

Similarly, the continuation of K'_{III} across the ray $\theta < -\omega$ into

$$-2\pi < \arg(\theta + \omega) < -\pi$$

$$\text{is} \quad -\int_0^\infty J_\nu(\nu t) e^{\nu\theta} d\nu = K'_{IV}. \quad (34)$$

Now by (20)_I,

$$K_{IV} = -\frac{1}{2} \int_0^\infty H_\nu^{(1)}(\nu t) e^{-\nu\theta} d\nu - \frac{1}{2} \int_0^\infty H_\nu^{(2)}(\nu t) e^{-i\nu\theta} d\nu,$$

where by (24), (25) the respective integrands are of orders $\nu^{-\frac{1}{2}} e^{i\nu(\omega-\theta)}$, $\nu^{-\frac{1}{2}} e^{-i\nu(\omega+\theta)}$. Hence if $-\omega < \theta < \omega$, we can rotate the paths to the rays $\arg \nu = \pi$, $-\pi$ respectively, which gives the continuation of K_{IV} across this segment as

$$\frac{1}{2} \int_0^\infty H_{-\lambda}^{(1)}(\lambda t e^{\pi i}) e^{i\lambda\theta} d\lambda + \frac{1}{2} \int_0^\infty H_{-\lambda}^{(2)}(\lambda t e^{-\pi i}) e^{i\lambda\theta} d\lambda = -\int_0^\infty J_\nu(\nu t) e^{\nu\theta} d\nu = K'_{IV}, \text{ by (34).}$$

Hence the continuation of $K_{III}(=K'_{III})$ in either sense round $\theta = -\omega$ leads to the same new branch $K_{IV}(=K'_{IV})$, and K has a simple branch point at $\theta = -\omega$.

(v) From the form of (33), (34) it is seen that K_{IV} , K'_{IV} must have the same sort of behaviour round $\theta = \omega$ as round $\theta = -\omega$, K (continued from K_{IV} or K'_{IV}) has a simple branch point at $\theta = \omega$, and the second branch for the neighbourhood of this point is

$$K_V = \int_0^\infty J_{-\nu}(\nu t) e^{-i\nu(\pi+\theta)} d\nu = \int_0^\infty J_{-\nu}(\nu t) e^{i\nu(\pi+\theta)} d\nu = K'_V. \quad (35)$$

Finally, K_V is regular in region VI as well as V; it can be continued thence into I; and in VI, I it is represented by the original series K_I .

5. ANALYTIC CONTINUATION OF $L_\mu(t, \theta)$ [EQUATION (16a)]

The procedure is exactly as in §4, and it will be sufficient to make a few remarks and to give the formulae.

(i) The case that ultimately concerns us is the one where μ is a zero of $J_\mu(\mu t)$; but it is convenient to suppose at first that μ is not real; for definiteness, $\text{Im } \mu > 0$.

(ii) The key formula is the analogue of (28):

$$\begin{aligned} \sum_1^\infty \frac{J_n(nt) e^{-in\theta}}{n-\mu} + \frac{\pi J_\mu(\mu t) e^{i\mu(\pi-\theta)}}{\sin \mu\pi} &= \frac{1}{2i} \int_{\infty e^{\pi i}}^{(0-)} \frac{J_\nu(\nu t) e^{i\nu(\pi-\theta)}}{(\nu-\mu) \sin \nu\pi} d\nu \\ &= \frac{1}{\mu} + \sum_1^\infty \frac{J_n(nt) e^{in\theta}}{n+\mu} - \int_0^\infty \frac{J_{-\nu}(\nu t) e^{-i\nu(\pi-\theta)} d\nu}{\nu+\mu}, \end{aligned}$$

which is valid for $t \geq 1$ and $\omega < \theta < 2\pi - \omega$. This shows that if we are to obtain continuation formulae resembling those of §4 we must start with the series

$$L_\mu = \sum_0^\infty \frac{J_n(nt) e^{in\theta}}{\mu+n} + \sum_1^\infty \frac{J_n(nt) e^{-in\theta}}{\mu-n} = L_{\mu I} = L_{\mu II}, \quad (36)$$

and not, say, with either half of it separately, and it gives the formula

$$L_{\mu\text{II}} = \int_0^\infty \frac{J_{-\nu}(\nu t) e^{-\nu(\pi-\theta)} d\nu}{\nu + \mu} + \frac{\pi J_\mu(\mu t) e^{i\mu(\pi-\theta)}}{\sin \mu\pi} = L_{\mu\text{III}}. \quad (37)$$

(iii) For the integral in (37), the rotation of the path of integration in the lower half-plane will give a passage across the pole $\nu = -\mu$. Hence we get terms in the formulae for $L'_{\mu\text{III}}$, etc., which have no counterparts in those for K'_{III} , etc.

(iv) The proof that L_μ has a simple branch point at $\theta = -\omega$ is just like that of §4. The formulae which give it in the four half-planes which make up the Riemann surface for this neighbourhood are (37), together with

$$\left. \begin{aligned} L'_{\mu\text{III}} &= \int_0^\infty \frac{J_{-\nu}(\nu t) e^{i\nu(\pi-\theta)} d\nu}{\mu - \nu} + \pi i H_\mu^{(1)}(\mu t) e^{i\mu(2\pi-\theta)} + \frac{\pi J_\mu(\mu t) e^{i\mu(\pi-\theta)}}{\sin \mu\pi}, \\ L_{\mu\text{IV}} &= \int_0^\infty \frac{J_\nu(\nu t) e^{-i\nu\theta} d\nu}{\nu - \mu} - 2\pi i J_\mu(\mu t) e^{-i\mu\theta} + \frac{\pi J_\mu(\mu t) e^{i\mu(\pi-\theta)}}{\sin \mu\pi}, \\ L'_{\mu\text{IV}} &= -\int_0^\infty \frac{J_\nu(\nu t) e^{i\nu\theta} d\nu}{\nu + \mu} - \pi i H_\mu^{(2)}(\mu t) e^{-i\mu\theta} + \frac{\pi J_\mu(\mu t) e^{i\mu(\pi-\theta)}}{\sin \mu\pi}. \end{aligned} \right\} \quad (38)$$

(v) Now let μ become a zero of $J_\mu(\mu t)$, as required for (16). In the formulae for $L_{\mu\text{III}}$, $L_{\mu\text{IV}}$ the terms in $J_\mu(\mu t)$ vanish and the integrands remain proper. But the formulae for $L'_{\mu\text{III}}$, $L'_{\mu\text{IV}}$ do not simplify in this way, and we shall not subsequently use them.

6. ANALYTIC CONTINUATION OF $\Omega^{(0)}(t, \theta)$

We define this function in region I by the expression on the left of (16):

$$\Omega^{(0)} = \Omega_I^{(0)} = \sum_0^\infty \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{\delta^n \Gamma(n+1)} + \sum_1^\infty \frac{h_n \delta^n \Gamma(n) \chi_n(\tau) e^{in\theta}}{(n/e)^n}. \quad (39)$$

Working at present formally, (16) gives

$$\Omega_I^{(0)} = QK_I - \sum_\mu f_\mu L_{\mu\text{I}},$$

and the continuations of $\Omega^{(0)}$ through regions II, III, ... are found by substituting for K , L_μ the appropriate continuations. Thus $\Omega_{\text{II}}^{(0)}$ is given by the series (39), while in region III

$$\begin{aligned} \Omega_{\text{III}}^{(0)} &= QK_{\text{III}} - \sum_\mu f_\mu L_{\mu\text{III}} \\ &= Q \int_0^\infty J_{-\nu}(\nu t) e^{-\nu(\pi-\theta)} d\nu - \sum_\mu f_\mu \int_0^\infty \frac{J_{-\nu}(\nu t) e^{-i\nu(\pi-\theta)} d\nu}{\nu + \mu} \end{aligned} \quad (40)$$

$$\begin{aligned} &= \int_0^\infty J_{-\nu}(\nu t) e^{-\nu(\pi-\theta)} \{Q - \sum_\mu f_\mu / (\nu + \mu)\} d\nu \\ &= \int_0^\infty \frac{\Gamma(\nu) \delta^\nu \sin \nu\pi \chi_{-\nu}(\tau) e^{-i\nu(\pi-\theta)} d\nu}{\pi(\nu/e)^\nu}, \end{aligned} \quad (41)$$

by (14) Similarly, in region IV we find, on use of (10),

$$\begin{aligned}\Omega_{IV}^{(0)} &= QK_{IV} - \sum_{\mu} f_{\mu} L_{\mu IV} \\ &= - \int_0^{\infty} J_{\nu}(\nu t) e^{-\nu\theta} \{Q + \sum f_{\mu}/(\nu - \mu)\} d\nu \\ &= - \int_0^{\infty} \frac{(\nu/e)^{\nu} \chi_{\nu}(\tau) e^{-\nu\theta} d\nu}{\delta^{\nu} \Gamma(\nu + 1)},\end{aligned}\quad (42)$$

and

$$\Omega_V^{(0)} = \int_0^{\infty} \frac{\Gamma(\nu) \delta^{\nu} \sin \nu\pi \chi_{-\nu}(\tau) e^{2\nu(\pi + \theta)} d\nu}{\pi(\nu/e)^{\nu}}. \quad (43)$$

Further, since K, L_{μ} have $\theta = -\omega$ as a simple branch point, it follows that $\Omega^{(0)}$ has a simple branch point, provided the series $\sum f_{\mu} L_{\mu}$ converges uniformly for θ near $-\omega$.

Note. These formulae for $\Omega^{(0)}$ may be established in the same way as we got the corresponding formulae for K in §4. But there appear to be no formulae for $\Omega^{(0)}$ of corresponding simplicity to those for K'_{III} and K'_{IV} , and to prove that $\Omega^{(0)}$ has a *simple* branch point at $\theta = -\omega$, it seems essential to proceed via the Kapteyn series.*

Justification. It is necessary (i) to justify the inversion of order of double summation (as in (16)) and summation and integration (as in (40)); (ii) to prove that the series $\sum f_{\mu} L_{\mu}$ remains uniformly convergent during the continuous variation of (t, θ) involved in the passage through the various regions.

As just noted, we can directly establish that the series (39) represents $\Omega^{(0)}$ in region II. Hence, so far as concerns the justifications (i) and (ii), it is sufficient to suppose $t > 1$.

Regarding (i), it will be sufficiently typical to examine the passage from (40) to (41). This is justified provided

$$\sum_{\mu} |f_{\mu}| \int_0^{\infty} \left| \frac{J_{-\nu}(\nu t) e^{-\nu(\pi - \theta)}}{\nu + \mu} \right| d\nu \quad (44)$$

exists. Consider a term for which $\mu = -\mu'$ is large negative. Then from (20)₂, (24), (25), and since $J_{-\mu'}(\mu' t) = 0$,

$$\int_0^{\infty} \left| \frac{J_{-\nu}(\nu t) e^{-i\nu(\pi - \theta)}}{\nu - \mu'} \right| d\nu \leq \int_0^{\mu'-1} \frac{2A d\nu}{\nu^{\frac{1}{2}}(\mu' - \nu)} + \int_{\mu'+1}^{\infty} \frac{2A d\nu}{\nu^{\frac{1}{2}}(\nu - \mu')} + \int_{\mu'-1}^{\mu'+1} \left| \frac{J_{-\nu}(\nu t) - J_{-\mu'}(\mu' t)}{\nu - \mu'} \right| d\nu.$$

The first two integrals on the right may be directly evaluated, and thereby proved to be $O(\mu'^{-\frac{1}{2}} \log \mu')$. The third is less than twice the greatest value in $(\mu' - 1, \mu' + 1)$ of

$$\left| \frac{\partial}{\partial \nu} J_{-\nu}(\nu t) \right|,$$

* To carry through on the lines of §4 we should need to use decompositions of χ_{ν} , $\chi_{-\nu}$ analogous to $J_{\nu} = \frac{1}{2}H_{\nu}^{(1)} + \frac{1}{2}H_{\nu}^{(2)}$. Such decompositions exist, but the components are not *completely* analogous to $H_{\nu}^{(1)}$, $H_{\nu}^{(2)}$; they have branch points and poles which greatly complicate the manipulations.

which is easily proved* to be $O(\mu'^{-1})$. Similarly, a term in (44) for which μ is large positive is proved to be $O(\mu^{-1})$, so the series (44) is dominated by

$$A'\Sigma\mu^{-1}|f_\mu| + A'\Sigma\mu'^{-1}\log\mu'|f_{-\mu'}|,$$

where μ runs through the positive zeros of $J_\nu(\nu t)$ and $-\mu'$ through the negative zeros. By (13) this is a convergent series.

Regarding (ii)—the uniform convergence of $\Sigma f_\mu L_\mu$ as t, θ vary during the continuation process—the essential point is that the cases in which ν, t, θ are real (of which a specimen has just been handled) are the *least* favourable ones; in general, the integrals that occur converge with exponential rapidity.

7. DERIVATION OF SOLUTIONS $\Omega^{(1)}, \Omega^{(2)}, \dots$

The fact that $\Omega^{(0)}$ is infinite on the branch lines $\theta = \pm\omega$ is easily verified from (16), (17); we find, for instance, for θ near $-\omega$,

$$\left. \begin{aligned} K_{\text{III}} &= \frac{(t^2-1)^{-\frac{1}{2}}}{2^{\frac{1}{2}}(\omega+\theta)^{\frac{1}{2}}} + O(1), & K_{\text{IV}} &= -\frac{(t^2-1)^{-\frac{1}{2}}}{2^{\frac{1}{2}}(\omega+\theta)^{\frac{1}{2}}} + O(1), \\ \Omega_{\text{III}}^{(0)} &= \frac{(1-\tau)^{\frac{1}{2}-\frac{1}{2}\beta}}{2^{\frac{1}{2}}(\tau/\tau_s-1)^{\frac{1}{2}}(\omega+\theta)^{\frac{1}{2}}} + O(1), \\ \Omega_{\text{IV}}^{(0)} &= -\frac{(1-\tau)^{\frac{1}{2}-\frac{1}{2}\beta}}{2^{\frac{1}{2}}(\tau/\tau_s-1)^{\frac{1}{2}}(\omega+\theta)^{\frac{1}{2}}} + O(1). \end{aligned} \right\} \quad (45)$$

We proceed therefore to obtain new solutions by integrating $\Omega^{(0)}$ any number of times with respect to θ .

Two points here need watching: (i) Formal θ -integration of (41) would lead to an integral divergent at $\nu = 0$, so we must first convert this to a form in which the path is detached from $\nu = 0$. (ii) We must verify that the integrated forms chosen in the several regions are continuations of each other, and do not differ by 'constants of integration' which here of course might be functions of τ .

The appropriate preliminary transformations of (41), (42) are

$$\Omega_{\text{III}}^{(0)} = 1 + \sum_1^\infty \frac{\Gamma(n) \delta^n h_n \chi_n(\tau) e^{in\theta}}{(n/e)^n} - \frac{1}{2i} \int_{\infty e^{-\pi i}}^{(0+)} \frac{(\nu/e)^\nu \chi_\nu(\tau) e^{i\nu(\pi-\theta)} d\nu}{\delta^\nu \Gamma(\nu+1) \sin \nu\pi}, \quad (46)$$

$$\Omega_{\text{IV}}^{(0)} = 1 + \sum_1^\infty \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{\delta^n \Gamma(n+1)} - \frac{1}{2i} \int_{\infty e^{-2\pi i}}^{(0+)} \frac{(\nu/e)^\nu \chi_\nu(\tau) e^{i\nu(\pi-\theta)} d\nu}{\delta^\nu \Gamma(\nu+1) \sin \nu\pi}; \quad (47)$$

in (46), $\arg \nu$ runs from $-\pi$ to π , while in (47) it runs from -2π to 0 , so that on the upper part of the path $(\nu/e)^\nu$ is not the principal value.

We now define

$$\Omega_{\text{III}}^{(r)} = \sum_1^\infty \frac{\Gamma(n) \delta^n h_n \chi_n(\tau) e^{in\theta}}{(in)^r (n/e)^n} - \frac{1}{2i} \int_{\infty e^{-\pi i}}^{(0+)} \frac{(\nu/e)^\nu \chi_\nu(\tau) e^{i\nu(\pi-\theta)} d\nu}{(-i\nu)^r \delta^\nu \Gamma(\nu+1) \sin \nu\pi}, \quad (48)$$

$$\Omega_{\text{IV}}^{(r)} = \sum_1^\infty \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{(-in)^r \delta^n \Gamma(n+1)} - \frac{1}{2i} \int_{\infty e^{-2\pi i}}^{(0+)} \frac{(\nu/e)^\nu \chi_\nu(\tau) e^{i\nu(\pi-\theta)} d\nu}{(-i\nu)^r \delta^\nu \Gamma(\nu+1) \sin \nu\pi}. \quad (49)$$

* Cauchy's integral gives $|\partial J_{-\nu}/\partial \nu|$ less than the greatest value of $2|J_{-\nu}(\nu t)|$ on the circle $|\nu - \mu'| = 2$; and by (24), (25) this is less than

$$4A(\mu' - 2)^{-\frac{1}{2}} e^{2(\pi+\omega)}.$$

To prove that $\Omega_{\text{III}}^{(r)}$, when continued round $\theta = -\omega$, has a simple branch point. Since $\Omega_{\text{III}}^{(1)}$ is a θ -integral of $\Omega_{\text{III}}^{(0)} - 1$, it can be continued along any path where $\Omega^{(0)}$ is regular; and since two circuits of θ round $-\omega$ restore the original determination of $\Omega_{\text{III}}^{(0)}$, they must lead $\Omega_{\text{III}}^{(1)}$ back to its original determination plus a constant C given by

$$C = \int (\Omega^{(0)} - 1) d\theta$$

round a curve which closes after two circuits. From (45) we can contract this double circuit on to the point $\theta = -\omega$, giving $C = 0$. Hence $\Omega^{(1)}$ has a simple branch point, and by repeating the argument we prove the same for $\Omega^{(2)}$.

To prove that $\Omega_{\text{IV}}^{(r)}$ is the second branch of $\Omega_{\text{III}}^{(r)}$. By (10), (24), (25) the integrand in (48) is of order $\nu^{-1} e^{w(2\pi-\theta-\omega)}$ in the upper half-plane, so for $\theta \leq 2\pi - \omega$ we can rotate the upper part of the path until it lies just below the positive half of the real axis. This gives

$$\begin{aligned} \Omega_{\text{III}}^{(r)} = & \sum_1^{\infty} \frac{\Gamma(n) \delta^n h_n \chi_n(\tau) e^{in\theta}}{(in)^r (n/e)^n} + \sum_1^{\infty} \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{(-in)^r \delta^n \Gamma(n+1)} \\ & - \frac{1}{2i} \int_{-\infty-i\epsilon}^{\infty-i\epsilon} \frac{(\nu/e)^{\nu} \chi_{\nu}(\tau) e^{w(\pi-\theta)} d\nu}{(-i\nu)^r \delta^{\nu} \Gamma(\nu+1) \sin \nu\pi}. \end{aligned} \quad (50)$$

Similarly, the integrand of (49) is, in the upper half-plane,* of order $\nu^{-\frac{1}{2}} e^{-w(\theta+\omega)}$, so for $\theta \leq -\omega$ we can rotate the upper part of the path until it lies just below the negative half of the real axis. This yields exactly the form on the right of (50), so $\Omega_{\text{III}}^{(r)} = \Omega_{\text{IV}}^{(r)}$ for $\theta \leq -\omega$, and $\Omega_{\text{IV}}^{(r)}$ must be the second branch of $\Omega_{\text{III}}^{(r)}$.

A formula resembling (48) for region V is easily established. Also for $\omega \leq \theta \leq 2\pi - \omega$, i.e. in region II, the integral in (50) is zero, since the path can be moved off to $-i\infty$, and this gives

$$\Omega_{\text{II}}^{(r)} = \Omega_{\text{I}}^{(r)} = \sum_1^{\infty} \frac{\Gamma(n) \delta^n h_n \chi_n(\tau) e^{in\theta}}{(in)^r (n/e)^n} + \sum_1^{\infty} \frac{(n/e)^n \chi_n(\tau) e^{-in\theta}}{(-in)^r \delta^n \Gamma(n+1)}. \quad (51)$$

The formulae (48), (49) can be put in 'real' forms resembling (41), (42) by reducing the path to the real axis and evaluating the contribution from the indentation round $\nu = 0$, which is not now zero. For example,

$$\begin{aligned} \Omega_{\text{IV}}^{(2)} = & \lim_{r \rightarrow 0} \left[\int_r^{\infty} \frac{(\nu/e)^{\nu} \chi_{\nu}(\tau) e^{-w\theta} d\nu}{\nu^2 \delta^{\nu} \Gamma(\nu+1)} - \frac{1}{r} + \frac{1}{2} \log^2 r + i\pi \log r - \frac{1}{6} \pi^2 \right. \\ & + (\log r + i\pi) \{c_1 + f_1 - \log(\delta e) - i(\pi + \theta)\} - \frac{1}{2}(\pi + \theta)^2 + \frac{1}{2} \log^2(\delta e) \\ & \left. + i(\pi + \theta) \log(\delta e) - (c_1 + f_1) \{i(\pi + \theta) + \log(\delta e)\} + c_2 + f_2 \right], \end{aligned} \quad (49a)$$

where the c_k, f_k are defined by the expansions, for $\nu \sim 0$,

$$1/\Gamma(\nu+1) = 1 + c_1 \nu + c_2 \nu^2 + \dots, \quad \chi_{\nu}(\tau) = 1 + \nu f_1(\tau) + \nu^2 f_2(\tau) + \dots$$

* Note that, before we can apply (10), we have to put

$$(\nu/e)^{\nu} \arg \nu = -2\pi = e^{-2\nu\pi i} (\nu/e)^{\nu} \arg \nu = 0.$$

8. THE FUNDAMENTAL PHYSICAL SOLUTION

We define this solution, P say, as minus the real part of $\Omega^{(2)}$, and shall prove that it satisfies the conditions laid down in §1 for flow of the Laval type. Thus from (51)

$$P_I = P_{II} = P_{VI} = \sum_1^\infty \left\{ \frac{\Gamma(n) \delta^n h_n}{(n/e)^n} + \frac{(n/e)^n}{\delta^n \Gamma(n+1)} \right\} \frac{\chi_n(\tau) \cos n\theta}{n^2}. \quad (52)$$

From (49a) we can pass to a formula for P_{IV} , which, however, we shall not need in the following discussion; and there are similar formulae for P_{III} , P_V .

The formulae for x , y are obtained from (2) by putting $\Omega = P$. Writing

$$X = 2\tau^{\frac{1}{2}} P_\tau, \quad Y = \tau^{-\frac{1}{2}} P_\theta, \quad (53)$$

we have

$$x/a = X \cos \theta - Y \sin \theta, \quad y/a = X \sin \theta + Y \cos \theta. \quad (54)$$

It is convenient to discuss X , Y in part from (53) and in part from (55) to (58) below.

From (52), (53) and (39), we have in region I

$$Y_\theta = \tau^{-\frac{1}{2}} P_{\theta\theta} = -\tau^{-\frac{1}{2}} \{\text{Rl } \Omega^{(0)} - 1\}, \quad (55)$$

$$X_\theta = 2\tau^{\frac{1}{2}} \frac{\partial}{\partial \tau} (\tau^{\frac{1}{2}} Y). \quad (56)$$

Also from (52), (53)

$$\tau^{\frac{1}{2}} X = \sum_1^\infty \left\{ \frac{\Gamma(n) \delta^n h_n}{(n/e)^n} + \frac{(n/e)^n}{\delta^n \Gamma(n+1)} \right\} \frac{\tau^{\frac{1}{2}} \chi'_n(\tau) \cos n\theta}{n^2}; \quad (57)$$

and since (see (67), appendix 1)

$$\frac{d}{d\tau} \{\tau^{\frac{1}{2}} \chi'_n(\tau)\} = \frac{1 - \tau/\tau_s}{2\tau^{\frac{1}{2}}(1-\tau)} \left(\frac{n^2}{2\tau} \chi_n(\tau) - \chi'_n(\tau) \right),$$

we get

$$X_\tau = \frac{1 - \tau/\tau_s}{2\tau^{\frac{1}{2}}(1-\tau)} \{\text{Rl } \Omega^{(0)} - 1 - \tau^{\frac{1}{2}} X\}. \quad (58)$$

By analytic continuation, (55), (56), (58) must be valid in all the regions.

Verification of the Laval conditions

(i) In regions I, IV, P is an even function of θ , so $\theta = 0$ gives $Y = 0$ and thence, by (54), $y = 0$. Hence the x -axis is a straight streamline, and the symmetry of the flow-field about it is verified.

(ii) We are to verify that, on the axis, the speed increases steadily with x , i.e. that x increases with τ , at least in a range surrounding τ_s . For $\tau < \tau_s$ this follows from (57), since then (appendix 1) each term $\tau^{\frac{1}{2}} \chi'_n(\tau)$ increases with τ (except for $n = 1$, when it is constant).

To discuss what happens for $\tau > \tau_s$ we appeal to (58), where for $\theta = 0$, X becomes x/a . Here $\Omega^{(0)} = QK - \Sigma f_\mu L_\mu$; by §6, $\Sigma f_\mu L_\mu$ is uniformly $O(1)$, and by (17), $K = (1-t)^{-1}$ for $\theta = 0$. Hence for τ near τ_s ,

$$\frac{x_\tau}{a} = \frac{Q}{2\tau^{\frac{1}{2}}(1-\tau)} \frac{1 - \tau/\tau_s}{1-t} + O(1 - \tau/\tau_s) + O\{x(1 - \tau/\tau_s)\},$$

where by (11) $(1 - \tau/\tau_s)/(1 - t)$ remains regular and positive as τ passes through τ_s . Hence it follows that x and x_τ remain regular and positive in some range $0 < \tau < \tau_1$, where $\tau_1 > \tau_s$. (It seems likely that $(1 - \tau/\tau_s) \text{Rl } \Omega^{(0)}$ is positive throughout $\tau_s < \tau < 1$, and if this is so x_τ also is positive throughout this range; to $\tau = 1$ corresponds, of course, the cavitation or limiting speed of the gas flow.)

(iii) Regarding the solution of (54) for τ, θ in terms of x, y , we obtain on the axis $\theta = 0$

$$\frac{1}{a} \frac{\partial(x, y)}{\partial(\tau, \theta)} = X_\tau(X + Y_\theta) = \frac{2\tau(1 - \tau) X_\tau^2}{\tau/\tau_s - 1},$$

where the second equality follows from (55), (58). Hence the proposed solution may be made, uniquely, in a strip bordering the axis $\theta = 0$ in region I, whose breadth tends to zero as τ approaches τ_s ; and similarly in region IV up to the value τ_1 (if any) at which x_τ vanishes. The same thing is established for the neighbourhood of the critical point $\tau = \tau_s, \theta = 0$ as follows:

In place of the variables τ, θ we use t, ϕ defined by (8) and (18); then the whole neighbourhood of the critical point is represented on a single-sheeted $t\phi$ -plane (figure 4), so X, Y are single-valued functions of t, ϕ , and the same formulae will serve for all the regions I to VI. Further, we obtain significant approximations from (55), (56), (58) by replacing $\text{Rl } \Omega^{(0)}$ by its principal part $QK = Q/(1 - t \cos \phi)$; and we keep only the leading terms, so that, for example, we replace Q by its value at the critical point Q_s given by (12). Thus (55) gives

$$\begin{aligned} \tau_s^{\frac{1}{2}} Y &= - \int_0^\theta \frac{Q_s d\theta}{1 - t \cos \phi} = - \int_0^\phi Q_s d\phi, \quad \text{from (18),} \\ &= - Q_s \phi. \end{aligned} \quad (59)$$

Thence from (56)

$$X_\theta = - \left(2\tau^{\frac{1}{2}} Q \frac{dt}{d\tau} \right)_s \frac{\partial \phi}{\partial t} = - \frac{Q_s}{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}} \frac{\sin \phi}{1 - t \cos \phi}, \quad \text{from (11), (18);}$$

from (58)

$$X_t = \frac{2Q_s}{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}} \frac{1 - t}{1 - t \cos \phi};$$

and thence

$$\begin{aligned} dX &= X_t dt + X_\theta d\theta = X_t dt + X_\theta \{(1 - t \cos \phi) d\phi - \sin \phi dt\} \\ &= \frac{Q_s}{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}} \left\{ \frac{2(1 - t) + \sin^2 \phi}{1 - t \cos \phi} dt - \sin \phi d\phi \right\}. \end{aligned}$$

Here, to leading order,

$$\frac{2(1 - t) + \sin^2 \phi}{1 - t \cos \phi} = \frac{1 - t^2 + t^2 \sin^2 \phi}{1 - t \cos \phi} = 1 + t \cos \phi = 2,$$

so the integration gives

$$X - X_s = \frac{Q_s}{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}} \{2(t - 1) - \frac{1}{2}\phi^2\}. \quad (60)$$

Let $t - 1$ be restricted to be $O(\phi^2)$. Then

$$\theta = \phi - t \sin \phi = \phi(1 - t + \frac{1}{6}\phi^2) = O(\phi^3), \quad (61)$$

and (54) give, to leading order,

$$\begin{aligned}(x - x_s)/a &= (X - X_s) + O(Y\theta) + O(\theta^2) = X - X_s = O(\phi^2), \\ y/a &= Y + O(\theta) + O(Y\theta^2) = Y = O(\phi).\end{aligned}$$

Thus the solution of (59), (60) is

$$\left. \begin{aligned} \frac{\tau - \tau_s}{2\tau_s(1 - \tau_s)^{\frac{1}{2}}} = t - 1 &= \frac{\tau_s}{4a^2Q_s^2}y^2 + \frac{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}}{2aQ_s}(x - x_s), \\ \theta &= \frac{\tau_s^{\frac{1}{2}}y}{aQ_s} \left\{ \frac{\tau_s}{12a^2Q_s^2}y^2 + \frac{\tau_s^{\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}}{2aQ_s}(x - x_s) \right\}. \end{aligned} \right\} \quad (62)$$

Hence τ and θ are regular functions of x, y near the critical point, as desired.

This completes the verification that the solution P specifies a flow of the Laval type.* Writing $C = 2a\tau_s^{-\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}}Q_s = 2a\tau_s^{-\frac{1}{2}}(1 - \tau_s)^{\frac{1}{2}(1-\beta)}$, the approximate equations of the loci in figure 1 are:

$$\text{The sonic line } aa': \quad t = 1. \quad y^2 = -C(x - x_s).$$

$$cc': \quad \theta = 0: \quad y^2 = -3C(x - x_s).$$

$$\text{The branch line } dd': \quad t - 1 = \frac{1}{2}\phi^2: \quad y^2 = C(x - x_s).$$

$$\text{The characteristic } bb': \quad \phi = -2\sqrt{2}\sqrt{(t-1)}: \quad y^2 = -2C(x - x_s).$$

9. OTHER PHYSICAL SOLUTIONS

A flow of Laval type is specified by the more general solution

$$\Omega = P + \sum_{\nu} A_{\nu} \chi_{\nu}(\tau) \cos \nu \theta, \quad (63)$$

where the added series is either finite, or infinite with $\sum |A_{\nu} \nu^2 \chi_{\nu}(\tau)|$ uniformly convergent; for the added series gives negligible contributions to the values of x_{τ} and $\partial(x, y)/\partial(\tau, \theta)$ near the critical point, e.g. their contribution to x_{τ} is zero at the critical point. There is here no restriction that ν be integral or even positive. By suitably choosing the A_{ν} , we can clearly adjust the velocity distribution on the axis.

The solutions

$$P_3 = \text{Im } \Omega^{(3)}, \quad P_4 = \text{Re } \Omega^{(4)}, \quad P_5 = \text{Im } \Omega^{(5)}, \quad \dots$$

have the correct symmetry and triple-valuedness (as functions of τ, θ) for a Laval-type flow, but each of them appears to give x_{τ} with a simple zero at the critical point; but a combination such that x_{τ} had a double zero at the critical point might be of interest. The solution

$$\Omega = P + \sum_3^{\infty} B_n P_n$$

will, of course, be of Laval type if the series is convergent with sufficient rapidity.

* The verification of condition (iv) of §1 is trivial.

APPENDIX 1

$\chi_\nu(\tau)$ is that solution of the differential equation

$$\chi_\nu'' + \left(\frac{1}{\tau} - \frac{\beta}{1-\tau} \right) \chi_\nu' + \nu^2 \left(\frac{\beta}{2\tau(1-\tau)} - \frac{1}{4\tau^2} \right) \chi_\nu = 0, \quad (64)$$

whose ratio to $\tau^{1/2}$ has the limit 1 as $\tau \rightarrow 0$. There is one such solution provided ν is not a negative integer, viz.

$$\chi_\nu(\tau) = \tau^{1/2} F(a_\nu, b_\nu; \nu + 1, \tau) = \tau^{1/2} (1 + \dots), \quad (65)$$

where $a_\nu + b_\nu = \nu + \beta$, $a_\nu b_\nu = -\frac{1}{2}\beta\nu(\nu - 1)$. As a function of ν , $\chi_\nu(\tau)$ has simple poles at the negative integers $\nu = -n$, with residues $-h_n \chi_n(\tau)$, where (Cherry 1947, §3)

$$h_n = \frac{(a_n - 1) \dots (a_n - n)(n - b_n) \dots (1 - b_n)}{n!(n-1)!} > 0 \quad (66)$$

From (64) we derive

$$\frac{d}{d\tau} (\tau^{1/2} \chi_\nu') = \frac{1 - \tau/\tau_s}{2\tau^{1/2}(1-\tau)} \left(\frac{\nu^2}{2\tau} \chi_\nu - \chi_\nu' \right), \quad (67)$$

$$\frac{d}{d\tau} \left(\frac{\tau \chi_\nu'}{\chi_\nu} \right) = \frac{\nu^2(1 - \tau/\tau_s)}{4\tau(1-\tau)} + \frac{\beta}{1-\tau} \frac{\tau \chi_\nu'}{\chi_\nu} - \frac{1}{\tau} \left(\frac{\tau \chi_\nu'}{\chi_\nu} \right)^2; \quad (68)$$

and from (65)

$$2\tau \chi_\nu' / \chi_\nu = \nu + O(\tau), \quad \text{for } \tau \sim 0. \quad (69)$$

To prove that, if $\nu > 1$, $\tau^{1/2} \chi_\nu'$ increases as τ increases from 0 to τ_s .

By (65) and (69), χ_ν begins increasing from 0, and $2\tau \chi_\nu' / \chi_\nu$ begins at a value, ν , less than ν^2 ; so by (67), $\tau^{1/2} \chi_\nu'$ begins increasing, and first becomes stationary when $\tau = \tau_s$ or when $2\tau \chi_\nu' / \chi_\nu = \nu^2$. But in the latter case, (68) gives

$$\frac{d}{d\tau} \left(\frac{\tau \chi_\nu'}{\chi_\nu} \right) = \frac{\nu^2(1 - \nu^2)}{4\tau} < 0,$$

so that $2\tau \chi_\nu' / \chi_\nu$ must decrease through the value ν^2 , and since it started at a value less than ν^2 , it cannot take the value ν^2 until after χ_ν has vanished. But χ_ν cannot vanish before χ_ν' , and χ_ν' cannot vanish before $\tau^{1/2} \chi_\nu'$ has stopped increasing. Hence $\tau^{1/2} \chi_\nu'$ first stops increasing at τ_s ; the order of events is (i) τ reaches τ_s , (ii) χ_ν' becomes negative, (iii) χ_ν becomes negative, (iv) $2\tau \chi_\nu' / \chi_\nu$ decreases through ν^2 , and $\tau^{1/2} \chi_\nu'$ becomes stationary for the second time.

APPENDIX 2. PARTIAL FRACTION SERIES OF BESSEL TYPE

Let $\Lambda_\nu(z)$ be defined by

$$J_\nu(z) = \frac{1}{\Gamma(\nu+1)} \left(\frac{z}{2} \right)^\nu \Lambda_\nu(z), \quad (70)$$

so that $\Lambda_\nu(z)$ is single-valued in both z and ν . As regards z it is even, and equal to $1 + O(z^2)$ for $z \sim 0$; as regards ν it has poles at $\nu = -1, -2, \dots$

From Debye's asymptotic formulae for $J_\nu(\nu t)$ —there are a number of cases, according to the values of t and $\arg \nu$ —we can deduce formulae for $\Lambda_\nu(\nu t)$; for example, if t is real and greater than 1, and $|\arg \nu| \leq \pi - \epsilon$,

$$\Lambda_\nu(\nu t) = \frac{2}{(t^2 - 1)^{\frac{1}{2}}} \left(\frac{2}{et} \right)^\nu \left\{ \cos(\nu\omega - \tfrac{1}{4}\pi) (1 + O(\nu^{-2})) + O(\nu^{-1} \sin(\nu\omega - \tfrac{1}{4}\pi)) \right\}, \quad (71)$$

where $\omega = \sqrt{(t^2 - 1)} - \arctan \sqrt{(t^2 - 1)}$. Now for $\tau_s < \tau < 1$ and $|\arg \nu| \leq \pi - \epsilon$ there is a similar formula for the hypergeometric function that occurs in (65).

$$\tau^{\frac{1}{2}} F(a_\nu, b_\nu; \nu + 1; \tau) = \frac{2\delta^\nu (1 - \tau)^{\frac{1}{2} - \frac{1}{2}\beta}}{(\tau/\tau_s - 1)^{\frac{1}{2}}} \left\{ \cos(\nu\omega' - \tfrac{1}{4}\pi) (1 + O(\nu^{-2})) + O(\nu^{-1} \sin(\nu\omega' - \tfrac{1}{4}\pi)) \right\}, \quad (72)$$

where
$$\omega' = \tau_s^{-\frac{1}{2}} \arctan \sqrt{\frac{\tau - \tau_s}{1 - \tau}} - \arctan \sqrt{\frac{\tau/\tau_s - 1}{1 - \tau}}.$$

By relating t to τ as in (8) we make $\omega' = \omega$, and obtain

$$\frac{F(a_\nu, b_\nu; \nu + 1; \tau)}{\Lambda_\nu(\nu t)} = (1 - \tau)^{\frac{1}{2} - \frac{1}{2}\beta} \left(\frac{t^2 - 1}{\tau/\tau_s - 1} \right)^{\frac{1}{2}} \left(\frac{\delta et}{2\sqrt{\tau}} \right)^\nu \left\{ 1 + O\left(\frac{1}{\nu}\right) \right\}, \quad (73)$$

provided ν is excluded from neighbourhoods of the zeroes of $\Lambda_\nu(\nu t)$; all that matters about these neighbourhoods is that their radii are small compared with the distance between neighbouring zeros.

By taking the analogues of (71), (72) for other values of τ and $\arg \nu$ we find the same result (73) in all cases, provided τ is not on $(1, +\infty)$. Of course, this treatment does not prove that the error term in (73) is uniformly $O(\nu^{-1})$ for τ near τ_s ; but this is true (see Cherry 1950), and the result is checked by the regularity of the factor

$$\{(t^2 - 1)/(\tau/\tau_s - 1)\}^{\frac{1}{2}}.$$

For shortness let (73) be written

$$\frac{F_\nu(\tau)}{S^\nu \Lambda_\nu(\nu t)} = Q \left\{ 1 + O\left(\frac{1}{\nu}\right) \right\}.$$

The function on the left (for τ fixed) is single-valued in ν , with simple poles at the zeros of $\Lambda_\nu(\nu t)$; the poles of $F_\nu(\tau)$ at $\nu = -1, -2, \dots$ are cancelled by those of $\Lambda_\nu(\nu t)$. Then

$$\frac{1}{2\pi i} \int_C \frac{F_z(\tau) dz}{S^z \Lambda_z(z t) (z - \nu)} = \frac{1}{2\pi i} \int_C \frac{Q dz}{z - \nu} \left\{ 1 + O\left(\frac{1}{z}\right) \right\},$$

provided the large circle C avoids the zeros of $\Lambda_z(z t)$. Evaluating the integral on the left by residues, and letting the radius of $C \rightarrow \infty$ we obtain

$$\frac{F_\nu(\tau)}{S^\nu \Lambda_\nu(\nu t)} = Q + \sum_\mu \frac{F_\mu(\tau)}{S^\mu (\nu - \mu) \partial \Lambda_\mu(\mu t) / \partial \mu}, \quad (74)$$

where $\mu = \mu(t)$ is the typical zero of $\Lambda_\mu(\mu t)$. This is equivalent to the form (10) given in §3, with

$$f_\mu(\tau) = S^{-\mu} F_\mu(\tau) \left/ \frac{\partial}{\partial \mu} \Lambda_\mu(\mu t) \right|. \quad (75)$$

The proof ensures the convergence of the series in (74), the mechanism of the convergence is that the large zeros of $\Lambda_\nu(\nu t)$ are near to zeros of $F_\nu(\tau)$, so that in

(75) $F_\mu(\tau)$ is small when μ is large. The actual estimation of $f_\mu(\tau)$ involves detailed work with asymptotic formulae which will not be given here. The results for $\tau_s < \tau < 1$ are given in (13). For $0 \leq \tau < \tau_s$ the n th zero is approximately

$$\mu = \mu_n = -n + e^{-2nu}/2\pi, \quad (76)$$

where $u = \text{arc tanh } \sqrt{(1-t^2)} - \sqrt{(1-t^2)}$, and here

$$f_{\mu_n}(\tau) = O(e^{-2nu}/n).$$

These estimates are not uniform up to $\tau = \tau_s$, but they are not far out. For $\tau = \tau_s$, $\mu_n = -n + \frac{1}{6}$ (approximately) and $f_{\mu_n}(\tau_s) = O(n^{-\frac{1}{6}})$, whereas (76) suggests

$$\mu_n = -n + (2\pi)^{-1}.$$

REFERENCES

- Cherry, T. M. 1947 *Proc. Roy. Soc. A*, **192**, 45.
 Cherry, T. M. 1950 *Trans. Amer. Math. Soc.* **68**, 224 (see especially §5.3).
 Lighthill, M. J. 1947a *Proc. Roy. Soc. A*, **191**, 323.
 Lighthill, M. J. 1947b *Proc. Roy. Soc. A*, **191**, 352.

The propagation of spherical blast

By G. B. WHITLAM, *Department of Mathematics, The University, Manchester*

(Communicated by S. Goldstein, F.R.S.—Received 11 January 1950—

Revised 29 June 1950)

The attenuation of spherical shocks at large distances from the origin is investigated mathematically. To do this a general theory is developed of spherical wave motion at large distances, where the linearized approximations of the theory of sound are valueless owing to the divergence of the characteristics at infinity. In the case of disturbances small from the outset, this general theory is used to modify the linearized approximation to give a theory which is uniformly valid at all distances from the origin. The equation of the shock is

$$a_0 t = r - b \log^{\frac{1}{2}} r - b_1 - b_2 \log^{-\frac{1}{2}} r + O(\log^{-1} r),$$

where r is the distance from the origin at time t , measured in multiples of some characteristic length, a_0 is the velocity of sound in the undisturbed air and b , b_1 , b_2 are arbitrary constants, of which b is readily calculated on the assumption of small disturbances. If a second shock is produced, the pressure at a point between the shocks falls approximately linearly with time at a rate

$$\{2\gamma/(\gamma+1)\} a_0 r^{-1} \log^{-1} r \text{ atm./sec.}$$

For a weak explosion, the modified linear theory is used to show that behind the front shock an envelope of characteristics is formed and hence a second shock must appear.

1. INTRODUCTION

The problems considered in this paper are those in which a spherical shock is expanding into still air with decreasing velocity and strength. The blast produced by an explosion will be of this type, and since an explosion is essentially the introduction

of a finite mass of fluid in a certain region, it may be represented as a solid sphere which expands suddenly into still air but slows down to rest at a finite distance from the origin (figure 1). Alternatively, an explosion may be represented as the sudden release of a spherical region of uniform high pressure, and this, too, may be investigated by the methods to be described. Similarly, the propagation of any progressive wave with decreasing amplitude may be treated.

The theory is analogous to that for supersonic flow past a pointed body of revolution (Whitham 1950). (The direct analogy to the body of revolution is of course that of unsteady motion with cylindrical symmetry, for which the results are closely similar, for example, the strength of an expanding cylindrical shock at a large distance r from the axis is proportional to r^{-1}). In figure 1, OAB is the 'piston curve' for an explosion, S is the front shock and S_1 is the second shock which is found to appear in the flow (§5). In the (r, t) plane the correspondence to supersonic flow past a body is clearly seen; the theory will be developed therefore with constant reference to the previous work.

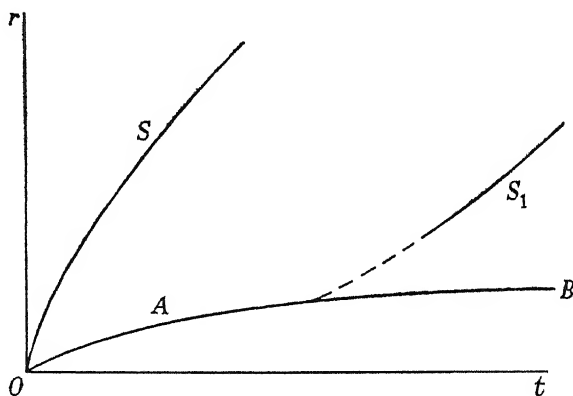


FIGURE 1.

As before, there is a linearized (acoustic) theory of the motion in which the characteristics are straight and parallel, being lines $r - a_0 t = \text{constant}$, where a_0 is the velocity of sound in the undisturbed air. Except for rather weak explosions, fluid velocities will be small enough (compared with that of sound) for this to be a good approximation only after a short time has elapsed, or at some distance from the origin. However, the approximation also becomes less and less valid at *large* distances from the origin, since the characteristics diverge (the wavelets near the head of the wave travelling slightly faster than those near the tail). The correct form at large distances is found in §2, where the exact equations of motion are solved using certain expansions for the velocity and the pressure. These results are used in §3 to find the shape of the front shock S in any problem for which there is also a second shock. For such a problem, the pressure variation with time at any point is approximately as shown in figure 2, hence it is called an *N*-wave. In §4, the general theory at large distances is used to modify the linearized approximation to give a theory which is uniformly valid at all distances from the origin, and in §5 the important case of an explosion is considered.

A numerical solution for the uniform expansion of a solid sphere has been found by Taylor (1946), and for the case when the velocity of the sphere is small compared with that of sound, Lighthill (1948) has given an analytic solution to the problem. It is found that a shock of uniform strength (expanding with constant velocity) is produced, and behind the shock, physical quantities are constant on lines $r/t = \text{constant}$. The problem is necessarily artificial in that the amplitude of the progressive wave is maintained to infinity, whereas ultimately the wave *must* decay, since the sphere must decelerate. Then, wavelets are emitted which interact with the shock producing a decrease of its strength, the attenuation is described in this paper.

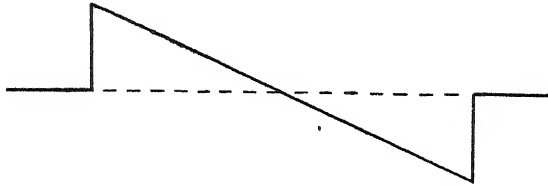


FIGURE 2.

Previous writers (Dumond, Cohen, Panofsky & Deeds 1946, Brinkley & Kirkwood 1947) have discussed the problem of an attenuating shock but stress that the assumptions of adiabatic flow are inadequate and that it is very necessary to take into account the entropy increment at the shock, that is, the degradation of energy into heat as the wave propagates. In the approach adopted here, it is agreed that this shedding of energy occurs, but the author thinks that the results can be obtained advantageously without explicit consideration of the phenomenon. The method is to find a solution of the equations of *isentropic* motion (valid at large distances from the source) to determine the flow behind the shock, and then the position and strength of the shock, required to give this flow, are determined from the Rankine-Hugoniot shock relations. It is felt that this method gives a clearer insight into the whole problem, fewer assumptions have to be made, and, in addition to finding the asymptotic equation of the shock, the flow behind the shock is described and the equations of the characteristics, which are important in themselves, are determined. Moreover, in the case when the explosion or motion of the sphere is small, the results are used to show the connexion of the flow at large distances with that near the source and, for example, with the motion of the sphere itself. This is obviously desirable, for instance, it enables the constants in the equation of the shock, which depend entirely on the initial motion, to be estimated.

The author is indebted to the referee of the paper for pointing out that, in its original form, the presentation was not convincing. The motion of the air is assumed to be isentropic, and the isentropic results are used to discuss the non-isentropic problem of an attenuating shock, some explanation is clearly required. The method is a general one and has been used in the cases of plane waves of one-dimensional unsteady flow (Friedrichs 1948), two-dimensional steady supersonic flow (Lighthill 1944) and supersonic flow past a body of revolution (Whitham 1950); these cases are included in the following discussion.

The isentropic equations of motion are used not because the flow is assumed to be *exactly* isentropic but because it is thought that the solution they give will be a good

approximation to the correct (non-isentropic) one, since the entropy changes at the shock are only of the third order in the strength. This is the only assumption,* and the results deduced from it are correct as far as is known, in that they agree with those obtained by energy considerations, in which the entropy changes are included explicitly. Hence there is some justification for the assumption, the 'isentropic' expressions for the velocity components do give an approximation to the correct ones, though it is difficult to estimate the accuracy of the results.

For the simpler case of plane waves, Lighthill has investigated the accuracy of the approach, in a paper to be published roughly simultaneously with this paper. A hypothesis on the accuracy of Friedrichs's work is made as follows 'The errors in the Friedrichs theory are comparable with the changes produced by altering the simple wave in two respects: (i) the specific entropy of the whole wave is altered to its known maximum, i.e. by an amount of order the cube of the maximum shock strength; (ii) the velocity distribution is altered, say by changing the velocity of the piston (continuously, and even after it has actually come to rest) by an amount whose ratio to the velocity of sound is of order the cube of the maximum shock strength.' This hypothesis is carefully examined by applying the theory to consideration of energy balance, and is substantiated by the checks made on it.

From this investigation, the Friedrichs theory is justified as a useful approximation, and therefore there is considerable value in an approach of the same type applied to the problem of spherical waves, in which the entropy changes are presumably of no greater importance than in the problem of plane waves. The accuracy expected is therefore the same as in the above hypothesis, modified to apply to spherical waves (e.g. for 'simple wave' read 'isentropic progressive wave').

2. EXPANSIONS OF THE PHYSICAL QUANTITIES FOR LARGE r

The equation of spherically symmetrical isentropic motion of the air is

$$a^2 \nabla^2 \phi = \frac{\partial^2 \phi}{\partial t^2} + 2 \frac{\partial \phi}{\partial r} \frac{\partial^2 \phi}{\partial r \partial t} + \left(\frac{\partial \phi}{\partial r} \right)^2 \frac{\partial^2 \phi}{\partial r^2}, \quad (1)$$

where ϕ is the velocity potential and a is the local velocity of sound given by Bernoulli's equation

$$a^2 = a_0^2 - (\gamma - 1) \left[\frac{\partial \phi}{\partial t} + \frac{1}{2} \left(\frac{\partial \phi}{\partial r} \right)^2 \right], \quad (2)$$

where a_0 is the velocity of sound in the undisturbed air. With $u = \partial \phi / \partial t$ and $v = \partial \phi / \partial r$, the equations of motion may be written

$$v_t - u_r = 0, \quad (3)$$

$$v_r [a_0^2 - (\gamma - 1) u - \frac{1}{2} (\gamma + 1) v^2] - u_t - 2v v_t + \frac{2v}{r} [a_0^2 - (\gamma - 1) u - \frac{1}{2} (\gamma - 1) v^2] = 0. \quad (4)$$

* Notice that the 'angle property' in § 3 is only quoted as a first order result, but the error term is in fact included, and taken account of in the subsequent work.

In the linearized theory, $\phi = f_0(a_0 t - r)/r$, hence u and v are of the form

$$u = \frac{f_1(a_0 t - r)}{r}, \quad (5)$$

$$v = \frac{f_2(a_0 t - r)}{r} + \frac{f_3(a_0 t - r)}{r^2}, \quad (6)$$

that is, u and v are expanded in negative powers of r with coefficients constant on each approximate characteristic $a_0 t - r = \text{constant}$. To obtain a solution uniformly valid for large r , expansions for u and v of a similar form are sought with coefficients constant on each *exact* characteristic $z = \text{constant}$, where z is a function of r and t to be determined in the process. Hence u and v are assumed to be of the form

$$u = a_0^2[f(z)r^{-1} + g(z)r^{-2} + \dots], \quad (7)$$

$$v = -u/a_0 + a_0[b(z)r^{-2} + c(z)r^{-3} + \dots]. \quad (8)$$

Substitution of these in the condition for $z = \text{constant}$ to be a characteristic ((10) below) suggests a similar expansion* for t :

$$a_0 t = r - z \log r - h(z) - m(z)r^{-1} - \dots \quad (9)$$

However, it is found that (7) and (8) require modification (with corresponding modification of (9)) in order to satisfy the equations of motion (3) and (4); to do this g, b, c, m in the above are replaced by

$$g_1(z) \log r + g_2(z), \quad b_1(z) \log r + b_2(z), \quad c_1(z) \log r + c_2(z), \quad m_1(z) \log r + m_2(z),$$

respectively. Now, substituting in (3) and (4) and equating coefficients, and satisfying the condition that $z = \text{constant}$ is a characteristic curve of the equations (3) and (4), that is along it,

$$\left(\frac{dt}{dr}\right)^2 [a_0^2 - (\gamma - 1)u - \frac{1}{2}(\gamma + 1)v^2] + 2v \frac{dt}{dr} - 1 = 0, \quad (10)$$

the unknown functions are determined in terms of $h(z)$ and certain arbitrary constants; $g_1(z)$ is found to be identically zero. The solution is

$$u = a_0^2 \left[-\frac{kz}{r} + \frac{K_1 z^2 + \frac{1}{2}B_1}{r^2} + \dots \right], \quad (11)$$

$$v = -\frac{u}{a_0} - a_0 \left[\frac{(\frac{1}{2}kz^2 + B_1) \log r + \frac{1}{2}kz^2 + k \int_0^z \zeta h'(\zeta) d\zeta + B_2}{r^2} \right] + \dots, \quad (12)$$

$$a_0 t = r - z \log r - h(z) - \frac{(\frac{1}{2}kz^2 + B_1) \log r + K_2 z^2 + \frac{1}{4}(\gamma + 5)B_1 + \int_0^z \zeta h'(\zeta) d\zeta + B_2}{r} + \dots, \quad (13)$$

where B_1 and B_2 are arbitrary constants, $k = 2/(\gamma + 1)$, $K_1 = k^2 - \frac{1}{4}k$, $K_2 = \frac{5}{4} + \frac{3}{2}k$.

* The coefficient of $\log r$ in the expansion of $r - a_0 t$ at large r has been taken equal to z itself rather than another function of z , since otherwise z will not be defined explicitly, being only required to be constant on the characteristics.

3. THE DECAY OF AN N -WAVE

The decay of an N -wave has been mentioned in §1 and refers here to any problem in which a spherical shock S , propagated outwards into still air, is followed by another shock S_1 . Due to interaction with the wavelets in the region between S and S_1 , S is retarded and S_1 accelerated; the strengths of S and S_1 tend to zero and their velocities to the sound velocity a_0 as $r \rightarrow \infty$.

There are two boundary conditions to be satisfied at a shock, and the most convenient forms for these are: (i) the 'angle property' which says that in the $(r, a_0 t)$ plane, to the first order in the strength of the shock, the angles that the shock makes with the characteristics on each side of it are equal, and (ii) ϕ is continuous at the shock, hence $\phi_r + U^{-1}\phi_t$ takes the same value on each side, where U is the velocity of the shock.

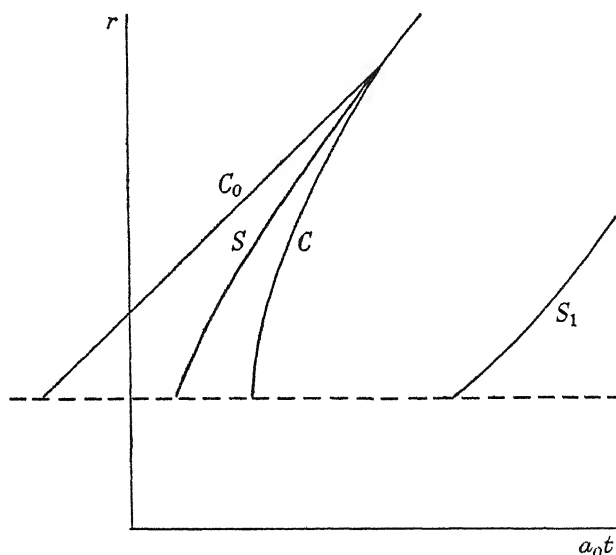


FIGURE 3.

Let C_0 (figure 3) be a characteristic of the undisturbed region ahead of S , and C a characteristic of the region between the shocks, then the equation of C is

$$a_0 t = r - z \log r - h(z) + O(r^{-1} \log r), \quad z = \text{constant}. \quad (14)$$

For *any* fixed r , t is bounded because the characteristic must lie between the two shocks S and S_1 , hence z and $h(z)$ are bounded in this region. Let the equation of S be $a_0 t = r - f(r)$, then from the angle property (i),

$$f'(r) = \frac{1}{2} z r^{-1} + O(r^{-2} \log r) \quad (15)$$

at the shock, and eliminating $a_0 t - r$ from the equations of C and S ,

$$f(r) = z \log r + h(z) + O(r^{-1} \log r). \quad (16)$$

The value of r at a point on the shock can be considered as a function of z , the value on the characteristic through that point of the shock, then, if this function $r(z)$ is

known, (16) gives f as a function of z and therefore the equation of S is determined with z as parameter. Differentiation of (16) with respect to z and substitution for $f'(r)$ from (15) gives

$$[zr^{-1} + O(r^{-2} \log r)] dr/dz + 2 \log r = -2h'(z),$$

or

$$d[z^2 \log r + O(zr^{-1} \log r)]/dz = -2zh'(z),$$

hence

$$z^2 \log r + O(zr^{-1} \log r) = -2 \int zh'(z) dz = -2zh(z) + 2h_1(z) + b^2,$$

where

$$h_1(z) = \int_0^z h(z) dz$$

and b is an arbitrary constant. Therefore

$$\log r = \frac{b^2}{z^2} - \frac{2h(z)}{z} + \frac{2h_1(z)}{z^2} + O(r^{-1} \log^{\frac{1}{2}} r), \quad (17)$$

and from (16)

$$f = \frac{b^2}{z} - h(z) + \frac{2h_1(z)}{z} + O(r^{-1} \log r). \quad (18)$$

Now, since $h(z)$ is bounded it may be expanded as $h(0) + zh'(0) + O(z^2)$, and eliminating z , the equation of S is

$$a_0 t = r - b \log^{\frac{1}{2}} r - h(0) - \frac{1}{2}bh'(0) \log^{-\frac{1}{2}} r + O(\log^{-1} r). \quad (19)$$

The error term in this is so large that it is not a very satisfactory form of the results, if the function $h(z)$ is known the equations (17) and (18), in which the error terms are of order r^{-1} with a logarithmic factor, may be used and will give a very much better approximation. The velocity U of the shock is the slope of S in the (r, t) plane, hence $1/U = (dt/dr) = (1 - f'(r))/a_0$, and condition (ii) gives $a_0 v + u - uf'(r) = 0$ immediately behind S , hence from (12) $B_1 = 0$, $B_2 = -\frac{1}{2}kb^2$.

For an N -wave if it were assumed that the flow behind the rear shock S_1 returns to the original undisturbed state, the equation of S_1 could be found by exactly the same arguments as for S to be

$$a_0 t = r + b_1 \log^{\frac{1}{2}} r + O(1), \quad (20)$$

where b_1 is an arbitrary constant. However, this assumption is not certainly a good approximation even at large distances from the origin. Therefore a discussion of the rear shock should include a disturbed state behind it; but then the same difficulties are encountered as in the case of the body of revolution, and similar conclusions may be drawn. They are not reconsidered here because nothing new can be added, and it may be found that, for practical applications to the effect of explosions, the assumptions on which equation (20) is based are adequate.

The excess pressure p at a point between the shocks is given by

$$-\rho_0 \partial \phi / \partial t + O[(\partial \phi / \partial r)^2],$$

that is

$$p/p_0 = \gamma kz/r + O(z^2 r^{-2}),$$

where p_0 and ρ_0 are the undisturbed pressure and density, and z is determined at the point (r, t) from (14). Expanding $h(z)$,

$$\frac{p}{p_0} = \frac{2\gamma}{\gamma+1} \frac{r - a_0 t - h(0)}{r \log r} + O(r^{-1} \log^{-\frac{3}{2}} r), \quad (21)$$

but, as before, if $h(z)$ is known a better approximation can be found. The pressure falls *linearly* with time at a fixed point between the shocks at a rate

$$\{2\gamma/(\gamma+1)\} a_0 r^{-1} \log^{-1} r,$$

which is independent of the mode of production of the shocks.

From (17), the value of z on a characteristic which meets the front shock at r is $O(\log^{-\frac{1}{2}} r)$; hence as $r \rightarrow \infty$, $z \rightarrow 0$. Therefore there exists a characteristic C^* , between the two shocks, which has the equation $a_0 t = r - h(0) + O(r^{-1} \log r)$; it is asymptotically a straight line and it does not meet either of the shocks. It may be observed that the arguments of this section would be valid for any flow in which some characteristic behind the front shock is asymptotically straight, for the presence of a rear shock was only used in proving that z and $h(z)$ are bounded on characteristics which meet the front shock, and the argument would still apply, since these characteristics must lie between the front shock and the straight characteristic. This will be used in § 5.

4. CONNEXION WITH LINEARIZED THEORY

In linearized theory, the distribution is assumed to be small and the solution of the equations is

$$u = \frac{a_0^2 F(\xi)}{r}, \quad (22)$$

$$v = -\frac{u}{a_0} - \frac{a_0 \int_0^\xi F(\xi') d\xi'}{r^2}, \quad (23)$$

where $\xi = a_0 t - r$ and the air is assumed to be undisturbed ahead of the characteristic $\xi = 0$. The arbitrary function $F(\xi)$ is determined from boundary conditions or initial values. If the disturbance is produced by the motion of a solid sphere, specified by $r = R(a_0 t)$ say, the boundary condition is that on $r = R(a_0 t)$, $\partial\phi/\partial r = v = dR/dt$; this is satisfied by

$$F(\xi) = -\frac{d}{d\xi} [R^2(\xi) R'(\xi)], \quad (24)$$

neglecting terms of the fifth order in R , R' and R'' .

The results of § 2, valid for large r , may be written in a form which compares directly with (22) and (23) by an appropriate change of the variable z . The arbitrary function $h(z)$ must contain a term of the form $-z \log d$, where d has the dimensions of a length, because $z \log r$ occurs in the series; choose $y = h(z) + z \log d$ and suppose z is then $-k^{-1}F(y)$. The results of § 2 become

$$u = a_0^2 \left[\frac{F(y)}{r} + \frac{\frac{1}{2} B_1 + O(F^2)}{r^2} + \dots \right], \quad (25)$$

$$v = -\frac{u}{a_0} - a_0 \left[\frac{(B_1 + O(F^2)) \log r + \int_{y_0}^y F(y') dy' + B_2 + O(F^2)}{r^2} \right] + \dots, \quad (26)$$

$$y = a_0 t - r - k^{-1} F(y) \log \frac{r}{d} - \dots, \quad (27)$$

where y_0 is the zero of $F(y)$ (i.e. the value on an asymptotically straight characteristic C^*). Equations (22) and (23) of linearized theory have the same form as (25) and (26) if squares of $F(y)$ are neglected and the arbitrary constants B_1 and B_2 are taken to be O and $\int_0^{y_0} F(y') dy'$ respectively. However, linearized theory fails at large distances because it uses the approximate characteristic variable ξ instead of the exact one y determined by (27). Hence, when the motion of the sphere is small so that by (24) F is small (of order δ^3 , where δa_0 is the maximum velocity of the sphere, say), a solution uniformly valid for *all* r is obtained by taking the linearized approximation with ξ replaced everywhere by y , and it is sufficient to determine y from $y = a_0 t - r - k^{-1} F(y) \log(r/d)$. Now, d has not been determined by this method, essentially because in this form of the results the variable y is only chosen to be constant on an exact characteristic and is not specified precisely. However, near the body the whole term $-k^{-1} F(y) \log(r/d)$ is $O(\delta^3)$ and may be neglected, at large distances, the exact expression for d , which must be of the order of the maximum radius of the sphere (the only fundamental length occurring), is found to be unnecessary as follows. In equation (17),

$$b^2 = -2k^{-1} B_2 = O(\delta^3) \quad \text{and} \quad z = -k^{-1} F(y) = O(\delta^3),$$

hence $\log r = O(\delta^{-3})$. Therefore, the equation deduced for the shock is only applicable for $\log r = O(\delta^{-3})$; again, equation (19) gives the strength of the shock, at a distance r to be proportional to $br^{-1} \log^{-\frac{1}{2}} r$, but the initial strength taken from the case of a uniformly expanding sphere (Lighthill 1948) is $O(e^{-(\text{constant})/\delta^3})$. Hence the theory only applies at distances r such that $\log r = O(\delta^{-3})$. For distances of this order $\log d$ is negligible in comparison with $\log r$.

Since the equation of the shock is not valid until distances of order e^{1/δ^3} are reached, the usefulness of the results of §3 may be doubted. However, one should observe that as the strength of the explosion (and hence δ) *increases*, the distance at which the theory may be applied *decreases*. Therefore, for strong explosions a good approximation should be obtained much earlier than suggested above. Physically, the explanation is that the attenuation of the shock is caused by the *accumulated* effect of sound convection, and this becomes appreciable over a smaller distance for a stronger explosion.

5. DISCUSSION OF AN EXPLOSION

The problem of an explosion will be investigated using the above modification of linearized theory. This holds strictly only for small motion of the sphere, but is hoped to give an adequate guide to the behaviour of strong explosions. In §3 the equation of the front shock has been found subject to the condition that there exists a straight characteristic in the flow, that this is so for any piston curve may be shown as follows. It is clear that either $R(\xi)$ has a maximum at a finite value ξ_0 , or $R'(\xi) > 0$ for all ξ and $R(\xi)$ tends to some fixed value R_0 as $\xi \rightarrow \infty$ (the latter being the case of an explosion). Hence $R^2(\xi) R'(\xi)$ is zero at $\xi = 0$, and is zero at ξ_0 or tends to zero as $\xi \rightarrow \infty$; in either case, by Rolle's theorem there exists a zero of $F(\xi) = -d(R^2 R')/d\xi$

other than $\xi = 0$. Therefore there exists a straight characteristic in the flow and the front shock is given by the results of §3, with

$$b^2 = -2k^{-1}B_2 = -2k^{-1} \int_0^{y_0} F(y') dy' = 2k^{-1}R^2(y_0) R'(y_0). \quad (28)$$

The question of whether a second shock is in fact formed may be investigated by discussing the formation of an envelope of characteristics. Such an envelope is given by equations

$$1 + k^{-1}F'(y) \log(r_1/d) = 0, \quad (29)$$

$$\alpha_0 t_1 = r_1 + k^{-1}F(y) \log(r_1/d) + y, \quad (30)$$

in which y is a parameter, and it exists for $F'(y) < 0$. For an explosion the form of $F(y)$ is shown in figure 4; hence envelopes could be formed by the characteristics

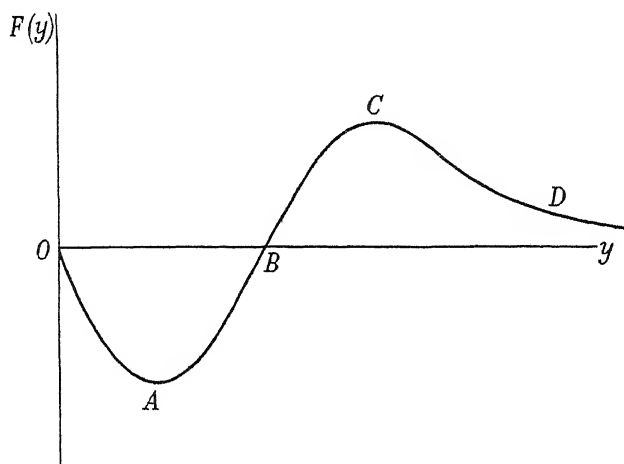


FIGURE 4.

corresponding to OA and CD . In OA , $F(y) < 0$; therefore from (30) the envelope would be a distance $O(\log r)$ ahead of the leading undisturbed characteristic whereas the shock is $O(\log^{\frac{1}{2}} r)$ ahead. Hence this envelope is not formed, the characteristics meet the shock before they meet one another. However, the characteristics from CD do actually form an envelope which indicates that a second shock will appear.

Example. Let the piston curve be $R(a_0 t) = \delta a_0 t / (1 + a_0 t)$ so that the initial velocity of the sphere is δa_0 and the ultimate radius is $\delta \times (\text{unit length})$. Then

$$F(y) = -2\delta^3 y(1-y)/(1+y)^5$$

(which is of the form shown in figure 4), $y_0 = 1$ and therefore

$$b^2 = 2k^{-1}R^2(1)R'(1) = 0.15\delta^3 \text{ for } \gamma = 1.4.$$

The excess pressure falls from the value $0.45\delta^{\frac{1}{2}}r^{-1}\log^{-\frac{1}{2}}r$ atm. at the front shock, at a rate $1.16a_0r^{-1}\log^{-1}r$ atm./sec. until the rear shock is reached, at which presumably it returns to a value small in comparison.

In conclusion, the representation of an explosion as the release of a spherical region of uniform high pressure P is considered. If c is the radius of the initial region, the linearized solution is (Jeffreys & Jeffreys 1946)

$$u = \frac{P}{2\rho_0 r} (a_0 t - r), \quad (31)$$

$$v = -\frac{u}{a_0} - \frac{P}{4\rho_0 a_0} \frac{[(a_0 t - r)^2 - c^2]}{r^2}, \quad (32)$$

for $|a_0 t - r| < c$ and $u = v = 0$ for $|a_0 t - r| > c$; it is already an N -wave. Modifying this by the above methods, a uniformly valid solution for all r is

$$u = \frac{P}{2\rho_0 r} \frac{y}{r}, \quad (33)$$

where
$$v = -\frac{u}{a_0} - \frac{P}{4\rho_0 a_0} \frac{(y^2 - c^2)}{r^2}, \quad (34)$$

$$y = a_0 t - r - \frac{P}{2\rho_0 a_0^2 k} y \log \frac{r}{\bar{a}}, \quad (35)$$

and in this case $B_2 = -Pc^2/4\rho_0 a_0^2$. The equation of the front shock is

$$a_0 t = r - b \log^{\frac{1}{2}} r + O(1),$$

where $b^2 = -2k^{-1}B_2 = (\gamma + 1)Pc^2/4\rho_0 a_0^2$, and there is strong indication that the rear shock is

$$a_0 t = r + b \log^{\frac{1}{2}} r + O(1). \quad (36)$$

To obtain the constant terms in the shock equations, higher approximations than the acoustic, to the motion near the origin, would presumably need to be known.

REFERENCES

- Brinkley, S. R. & Kirkwood, J. G. 1947 *Phys. Rev.* **71**, 606.
 Dumond, J. W. M., Cohen, E. R., Panofsky, W. K. H. & Deeds, E. 1946 *J. Acoust. Soc. Amer.* **18**, 97.
 Friedrichs, K. O. 1948 *Commun. Appl. Math.* **1**, 211.
 Jeffreys, H. & Jeffreys, B. S. 1946 *Methods of mathematical physics*, pp. 528–529. Cambridge University Press.
 Lighthill, M. J. 1944 *Rep. Memor. Aero. Res. Comm., Lond.*, no. 1930.
 Lighthill, M. J. 1948 *Quart. J. Mech. Appl. Math.* **1**, 309.
 Taylor, G. I. 1946 *Proc. Roy. Soc. A*, **186**, 273.
 Whitham, G. B. 1950 *Proc. Roy. Soc. A*, **201**, 89.

The application of phase-contrast to the ultra-violet microscope

BY E. W. TAYLOR, *Cooke, Troughton and Simms Ltd., York*

(Communicated by Sir Thomas Merton, *Treas.R.S.*—Received 1 February 1950)

(This paper has been printed in full in *Proceedings B*, 137, 332.)

ABSTRACT

It has been shown that the phase-contrast of Zernike may be applied with advantage to ultra-violet microscopy and that certain advantages follow:

(a) In the first place, it furnishes a more certain method than dark-ground illumination for the visual selection of a suitable field of view, as all structures are more clearly revealed.

(b) In the second place, it furnishes a new means whereby structural details that give rise to a change of phase in the transmitted radiations may be photographed in contrast, even under conditions when selective absorption does not take place.

(c) In the third place, it is possible, after having selected a suitable field by means of visual phase-contrast, to photograph the same field in ultra-violet light and then to turn once again to the visual phase-contrast image to make certain that the exposure has not caused damage to a living organism

A series of explanatory photographs has been taken covering a wide range of biological objects. These show that the method is sensitive to minute changes in phase, and that the resulting images are characterized by good contrast. Details are revealed that cannot be brought out by normal methods, as will be clear from a close study of the pairs of photographs which accompany this paper. In selecting these it was thought necessary to avoid very fine detail and delicate shades of contrast that would be lost in the process of reproduction.

INDEX TO VOLUME 203 (A)

- Adsorption of surface-active agents (Salley *et al.*), 42.
- Alloy formation (Frost & Raynor), 132.
- Argyle, Ann A. *See* Salley, Weith, Argyle & Dixon.
- Arlman, E. J. & Melville, H. W. Studies in copolymerization The evaluation of the kinetic coefficients for the system styrene-butyl acrylate, 301.
- Ashmore, P. G. & Norrish, R. G. W. A study of sensitized explosions. VIII. Experimental work on the hydrogen-oxygen reaction sensitized by chloropicrin, 454
- Ashmore, P. G. & Norrish, R. G. W. A study of sensitized explosions. IX. Experimental work on the sensitization of ignitions of hydrogen-chlorine mixtures by chloropicrin, 472.
- A study of sensitized explosions. VIII. Experimental work on the hydrogen-oxygen reaction sensitized by chloropicrin (Ashmore & Norrish), 454.
- A study in sensitized explosions. IX. Experimental work on the sensitization of ignitions of hydrogen-chlorine mixtures by chloropicrin (Ashmore & Norrish), 472.
- Athermal binary solutions, entropies of mixing (Guggenheim & McGlashan), 435.
- Atkins, K. R. Liquid helium films. I. The thickness of the film, 119
- Atkins, K. R. Liquid helium films. II. The flow of the film, 240.
- Atmospheric turbulence, fine structure (Richardson), 149.
- Atomic theory of elasticity (Huang), 178.
- Axisymmetric turbulence, decay (Chandrasekhar), 358.
- Bakerian Lecture. Physics above 20,000 kg/cm.² (Bridgman), 1.
- Bell, R. P. & Long, D. A. Polarizability and internuclear distance in the hydrogen molecule and molecule-ion, 364
- Bell Telephone Laboratories—an example of an institute of creative technology (Kelly), 287.
- Booth, F. The cataphoresis of spherical, solid non-conducting particles in a symmetrical electrolyte, 514.
- Booth, F. The electroviscous effect for suspensions of solid spherical particles, 533
- Bose-Einstein condensation (de Groot, Hooyman & ten Seldam), 266
- Bragg, Sir Lawrence, Kendrew, J. C. & Perutz, M. F. Polypeptide chain configurations in crystalline proteins, 321.
- Bridgman, P. W. Physics above 20,000 kg./cm.² (Bakerian Lecture), 1
- Carman, P. C. Diffusion and flow of gases and vapours through micropores. I. Slip flow and molecular streaming, 55.
- Carman, P. C. & Malherbe, P. le R. Diffusion and flow of gases and vapours through micropores. II. Surface flow, 165.
- Chandrasekhar, S. The decay of axisymmetric turbulence, 358.
- Cherry, T. M. Exact solutions for flow of a perfect gas in a two-dimensional Laval nozzle, 551.
- Chester, W. The propagation of a sound pulse in the presence of a semi-infinite, open-ended channel. II. 33.
- Conductivity of metals (MacDonald & Sarginson), 223.
- Copolymerization studies (Arlman & Melville), 301.
- Decay of axisymmetric turbulence (Chandrasekhar), 358.
- de Groot, S. R., Hooyman, G. J. & ten Seldam, C. A. On the Bose-Einstein condensation, 266.
- Derivation of a model solar chromosphere from radio data (Piddington), 417.

- Diffusion and flow of gases and vapours through micropores. I. Slip flow and molecular streaming (Carman), 55.
- Diffusion and flow of gases and vapours through micropores. II. Surface flow (Carman & Malherbe), 165.
- Dixon, J. K. *See* Salley, Weith, Argyle & Dixon.
- Eigenfunction problems with periodic potentials (Titchmarsh), 501.
- Elasticity, atomic theory (Huang), 178.
- Emission of heavy fragments in nuclear explosions (Perkins), 399.
- Entropies of mixing in certain athermal binary solutions (Guggenheim & McGlashan), 435.
- Exact solutions for flow of a perfect gas in a two-dimensional Laval nozzle (Cherry), 551
- Field variation of the superconducting penetration depth (Pippard), 210.
- Fine structure of atmospheric turbulence in relation to the propagation of sound over the ground (Richardson,) 149.
- Flow of a perfect gas (Cherry), 551.
- Frost, B. R. T. & Raynor, G. V. The system silver-magnesium-antimony, with reference to the theory of alloy formation, 132.
- Garrett, C. G. B. The nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K. I. Thermodynamic measurements made from a study of the field-dependence of the adiabatic susceptibility, 375.
- Garrett, C. G. B. The nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K. II. The nuclear specific heat, 392
- Guggenheim, E. A. & McGlashan, M. L. Entropies of mixing in certain athermal binary solutions, 435.
- Guggenheim, E. A. & Wiseman, L. A. Kinetic salt effects on the inversion of sucrose, 17.
- Helium, liquid films (Atkins), 119, 240.
- Hinshelwood, Sir Cyril. *See* Ingold, Stubbs & Hinshelwood.
- Hooyman, G. J. *See* de Groot, Hooyman & ten Seldam
- Huang, K. On the atomic theory of elasticity, 178.
- Ingold, K. U., Stubbs, F. J. & Hinshelwood, Sir Cyril. The kinetics of the thermal decomposition of normal paraffin hydrocarbons. III. Activation energies and possible mechanisms of molecular reactions, 486.
- Kelly, M. J. The Bell Telephone Laboratories—an example of an institute of creative technology, 287.
- Kendrew, J. C. *See* Bragg, Kendrew & Perutz.
- Kinetic salt effects on the inversion of sucrose (Guggenheim & Wiseman), 17
- Liquid helium films. I. The thickness of the film (Atkins), 119.
- Liquid helium films. II. The flow of the film (Atkins), 240.
- Long, D. A. *See* Bell & Long.
- MacDonald, D. K. C. & Sarginson, K. Size effect variation of the electrical conductivity of metals, 223.
- Malherbe, P. le R. *See* Carman & Malherbe.
- McGlashan, M. L. *See* Guggenheim & McGlashan.
- Measurement of the adsorption of surface-active agents at a solution/air interface by a radiotracer method (Salley, Weith, Argyle & Dixon), 42.
- Meksyn, D. Stability of viscous flow over concave cylindrical surfaces, 253.
- Melville, H. W. *See* Arlman & Melville.
- Micropores, flow of gases through (Carman), 55.
- Micropores, flow of gases through (Carman & Malherbe), 165.

- Norrish, R. G. W. *See* Ashmore & Norrish.
- Nuclear explosions, emission of heavy fragments (Perkins), 399
- Nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K (Garrett)
- I. Thermodynamic measurements made from a study of the field-dependence of the adiabatic susceptibility, 375.
- Nuclear specific heat in paramagnetic cupric salts at temperatures below 1° K (Garrett).
- II. The nuclear specific heat, 392.
- Paraffin hydrocarbons, thermal decomposition (Ingold, Stubbs & Hinshelwood), 486.
- Paramagnetic cupric salts, nuclear specific heat (Garrett), 375, 392.
- Perkins, D. H. Emission of heavy fragments in nuclear explosions, 399.
- Perutz, M. F. *See* Bragg, Kendrew & Perutz.
- Physics above 20,000 kg./cm.² (Bakerian Lecture), (Bridgman), 1.
- Piddington, J. H. The derivation of a model solar chromosphere from radio data, 417.
- Pippard, A. B. Field variation of the superconducting penetration depth, 210.
- Pippard, A. B. The surface impedance of superconductors and normal metals at high frequencies. IV. Impedance at 9400 Mc/sec of single crystals of normal and superconducting tin, 98.
- Pippard, A. B. The surface impedance of superconductors and normal metals at high frequencies. V. Analysis of experimental results for superconducting tin, 195.
- Polarizability and internuclear distance in the hydrogen molecule and molecule-ion (Bell & Long), 364.
- Polypeptide chain configurations in crystalline proteins (Bragg, Kendrew & Perutz), 321.
- Propagation of a sound pulse in the presence of a semi-infinite, open-ended channel.
- II (Chester), 33.
- Raynor, G. V. *See* Frost & Raynor.
- Richardson, E. G. The fine structure of atmospheric turbulence in relation to the propagation of sound over the ground, 149.
- Salley, D. J., Weith, A. J., Argyle, Ann A & Dixon, J. K. Measurement of the adsorption of surface-active agents at a solution/air interface by a radiotracer method, 42.
- Sarginson, K. *See* MacDonald & Sarginson.
- Sensitized explosions. VIII (Ashmore & Norrish), 454.
- Sensitized explosions. IX (Ashmore & Norris), 472.
- Size effect variation of the electrical conductivity of metals (MacDonald & Sarginson), 223.
- Solar chromosphere, model (Piddington), 417.
- Sondheimer, E. H. The theory of the transport phenomena in metals, 75.
- Sound pulse, propagation (Chester), 33.
- Spherical blast, propagation (Whitham), 571.
- Stability of viscous flow over concave cylindrical surfaces (Meksyn), 253.
- Stubbs, F. J. *See* Ingold, Stubbs & Hinshelwood.
- Studies in copolymerization. The evaluation of the kinetic coefficients for the system styrene-butyl acrylate (Arlman & Melville), 301.
- Styrene-butyl acrylate system (Arlman & Melville), 301.
- Sucrose inversion, kinetic salt effects on (Guggenheim & Wiseman), 17.
- Superconducting penetration depth (Pippard), 210.
- Superconductors, surface impedance (Pippard), 98, 195.
- Surface impedance of superconductors and normal metals at high frequencies. IV. Impedance at 9400 Mc/sec of single crystals of normal and superconducting tin, (Pippard), 98.
- Surface impedance of superconductors and normal metals at high frequencies. V. Analysis of experimental results for superconducting tin (Pippard), 195.
- System silver-magnesium-antimony, with reference to the theory of alloy formation (Frost & Raynor), 132.

- Taylor, E. W. The application of phase-contrast to the ultra-violet microscope (Abstract), 582.
- ten Seldam, C. A. *See* de Groot, Hooyman & ten Seldam.
- The application of phase-contrast to the ultra-violet microscope (Abstract) (Taylor), 582
- The cataphoresis of spherical, solid non-conducting particles in a symmetrical electrolyte (Booth), 514.
- The electroviscous effect for suspensions of solid spherical particles (Booth), 533
- The kinetics of the thermal decomposition of normal paraffin hydrocarbons III. Activation energies and possible mechanisms of molecular reactions (Ingold, Stubbs & Hinshelwood), 486.
- Theory of the transport phenomena in metals (Sondheimer), 75.
- The propagation of spherical blast (Whitham), 571
- Titchmarsh, E. C. Eigenfunction problems with periodic potentials, 501.
- Transport phenomena in metals (Sondheimer), 75.
- Viscous flow over concave cylindrical surfaces (Meksyn), 253.
- Weith, A. J. *See* Salley, Weith, Argyle & Dixon
- Whitham, G. B. The propagation of spherical blast, 571.
- Wiseman, L. A. *See* Guggenheim & Wiseman.

I A.R.I. 75

INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY, NEW DELHI.

[illegible]

GIPNLK—H-40 I.A.R.I.—29-4-55—15,000